

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 9/00		A2	(11) International Publication Number: WO 00/20566
			(43) International Publication Date: 13 April 2000 (13.04.00)
(21) International Application Number: PCT/US99/20797		Parkman Place, Cincinnati, OH 45213 (US). GLEESON, Martin [-/US]; 4228 Cordobes Cove, San Diego, CA 92130 (US).	
(22) International Filing Date: 10 September 1999 (10.09.99)			
(30) Priority Data: 60/103,099 5 October 1998 (05.10.98) US 60/123,555 10 March 1999 (10.03.99) US		(74) Agents: DRACH, John, E. et al.; Henkel Corporation, Patent Dept., Suite 200, 2500 Renaissance Boulevard, Gulph Mills, PA 19406 (US).	
(71) Applicant (for all designated States except US): HENKEL CORPORATION [US/US]; Suite 200, 2500 Renaissance Boulevard, Gulph Mills, PA 19406 (US).		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(72) Inventors; and (75) Inventors/Applicants (for US only): WILSON, C., Ron [-/US]; 6327 Belmont Road, Loveland, OH 45140 (US). CRAFT, David, L. [-/US]; 26 Rosewood Lane, Fort Thomas, KY 41075 (US). EIRICH, L., Dudley [-/US]; 227 W. Stoneridge Drive, Cincinnati, OH 45150 (US). ESHOO, Mark [-/US]; Apartment 5, 2821 Hillegass Avenue, Berkeley, CA 94705 (US). MADDURI, Krishna, M. [-/US]; Apartment C, 8237 Lenox Lane, Indianapolis, IN 46268 (US). CORNETT, Cathy, A. [-/US]; Apartment #11, 697 Meadow Wood Drive, Crescent Springs, KY 41017 (US). BRENNER, Alfred, A. [-/US]; 1563 Gamay Street, Santa Rosa, CA 95403 (US). TANG, Maria [-/US]; 2043 Cliffwood Drive, Fairfield, CA 94533 (US). LOPER, John, C. [-/US]; 6315		Published Without international search report and to be republished upon receipt of that report.	
(54) Title: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450 OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS RELATING THERETO			
(57) Abstract <p>Novel genes have been isolated which encode cytochrome P450 and NADPH reductase enzyme of the ω-hydroxylase complex of <i>C. tropicalis</i> 20336. Vectors including these genes, transfected host cells and transformed host cells are provided. Methods of producing of cytochrome P450 and NADPH reductase enzymes are also provided which involve transforming a host cell with a gene encoding these enzymes and culturing the cells. Methods of increasing the production of a dicarboxylic acid and methods of increasing production of the aforementioned enzymes are also provided which involve increasing in the host cell the number of genes encoding these enzymes. A method for discriminating members of a gene family by quantifying the expression of genes is also provided.</p>			

5 **CYTOCHROME P450 MONOOXYGENASE AND
NADPH CYTOCHROME P450 OXIDOREDUCTASE GENES AND
PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF
CANDIDA TROPICALIS AND METHODS RELATING THERETO**

CROSS REFERENCE TO RELATED APPLICATIONS

10 This application claims priority to U.S. Provisional Application Serial No.
60/103,099 filed October 5, 1998, and U.S. Provisional Application Serial No. 60/083,798 filed
May 1, 1998.

BACKGROUND

15 1. **Field of the Invention**

 The present invention relates to novel genes which encode enzymes of the ω -
hydroxylase complex in yeast *Candida tropicalis* strains. In particular, the invention relates to
novel genes encoding the cytochrome P450 and NADPH reductase enzymes of the ω -
hydroxylase complex in yeast *Candida tropicalis*, and to a method of quantitating the expression
20 of genes.

 2. **Description of the Related Art**

 Aliphatic dioic acids are versatile chemical intermediates useful as raw
materials for the preparation of perfumes, polymers, adhesives and macrolid antibiotics. While
25 several chemical routes to the synthesis of long-chain alpha, ω -dicarboxylic acids are available,
the synthesis is not easy and most methods result in mixtures containing shorter chain lengths.
As a result, extensive purification steps are necessary. While it is known that long-chain dioic
acids can also be produced by microbial transformation of alkanes, fatty acids or esters thereof,
chemical synthesis has remained the most commercially viable route, due to limitations with the
30 current biological approaches.

 Several strains of yeast are known to excrete alpha, ω -dicarboxylic acids as a
byproduct when cultured on alkanes or fatty acids as the carbon source. In particular, yeast
belonging to the Genus *Candida*, such as *C. albicans*, *C. cloacae*, *C. guilliermondii*, *C.*
intermedia, *C. lipolytica*, *C. maltosa*, *C. parapsilosis* and *C. zeylenoides* are known to produce

such dicarboxylic acids (*Agr. Biol. Chem.* 35: 2033-2042 (1971)). Also, various strains of *C. tropicalis* are known to produce dicarboxylic acids ranging in chain lengths from C₁₁ through C₁₈ (Okino et al., BM Lawrence, BD Mookherjee and BJ Willis (eds), in *Flavors and Fragrances: A World Perspective*. Proceedings of the 10th International Conference of Essential Oils, Flavors and Fragrances, Elsevier Science Publishers BV Amsterdam (1988)), and are the basis of several patents as reviewed by Bühler and Schindler, in *Aliphatic Hydrocarbons in Biotechnology*, H. J. Rehm and G. Reed (eds), Vol. 169, Verlag Chemie, Weinheim (1984).

Studies of the biochemical processes by which yeasts metabolize alkanes and fatty acids have revealed three types of oxidation reactions: α -oxidation of alkanes to alcohols, ω -oxidation of fatty acids to alpha, ω -dicarboxylic acids and the degradative β -oxidation of fatty acids to CO₂ and water. The first two types of oxidations are catalyzed by microsomal enzymes while the last type takes place in the peroxisomes. In *C. tropicalis*, the first step in the ω -oxidation pathway is catalyzed by a membrane-bound enzyme complex (ω -hydroxylase complex) including a cytochrome P450 monooxygenase and a NADPH cytochrome reductase.

15 This hydroxylase complex is responsible for the primary oxidation of the terminal methyl group in alkanes and fatty acids (Gilewicz et al., *Can. J. Microbiol.* 25:201 (1979)). The genes which encode the cytochrome P450 and NADPH reductase components of the complex have previously been identified as P450ALK and P450RED respectively, and have also been cloned and sequenced (Sanglard et al., *Gene* 76:121-136 (1989)). P450ALK has also been designated

20 P450ALK1. More recently, ALK genes have been designated by the symbol CYP and RED genes have been designated by the symbol CPR. See, e.g., Nelson, *Pharmacogenetics* 6(1):1-42 (1996), which is incorporated herein by reference. See also Ohkuma et al., *DNA and Cell Biology* 14:163-173 (1995), Seghezzi et al., *DNA and Cell Biology*, 11:767-780 (1992) and Kargel et al., *Yeast* 12:333-348 (1996), each incorporated herein by reference. For example,

25 P450ALK is also designated CYP52 according to the nomenclature of Nelson, *supra*. Fatty acids are ultimately formed from alkanes after two additional oxidation steps, catalyzed by alcohol oxidase (Kemp et al., *Appl. Microbiol. and Biotechnol.* 28: 370-374 (1988)) and aldehyde dehydrogenase. The fatty acids can be further oxidized through the same or similar pathway to the corresponding dicarboxylic acid. The ω -oxidation of fatty acids proceeds via the ω -hydroxy

30 fatty acid and its aldehyde derivative, to the corresponding dicarboxylic acid without the requirement for CoA activation. However, both fatty acids and dicarboxylic acids can be

- degraded, after activation to the corresponding acyl-CoA ester through the β -oxidation pathway in the peroxisomes, leading to chain shortening. In mammalian systems, both fatty acid and dicarboxylic acid products of ω -oxidation are activated to their CoA-esters at equal rates and are substrates for both mitochondrial and peroxisomal β -oxidation (*J. Biochem.*, 102:225-234 (1987)). In yeast, β -oxidation takes place solely in the peroxisomes (*Agr. Biol. Chem.* 49:1821-1828 (1985)).

- The production of dicarboxylic acids by fermentation of unsaturated C_{14} - C_{16} monocarboxylic acids using a strain of the species *C. tropicalis* is disclosed in U.S. Patent 4,474,882. The unsaturated dicarboxylic acids correspond to the starting materials in the number and position of the double bonds. Similar processes in which other special microorganisms are used are described in U.S. Patents 3,975,234 and 4,339,536, in British Patent Specification 1,405,026 and in German Patent Publications 21 64 626, 28 53 847, 29 37 292, 29 51 177, and 21 40 133.

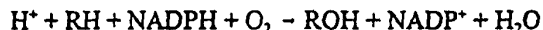
- Cytochromes P450 (P450s) are terminal monooxidases of a multicomponent enzyme system as described above. They comprise a superfamily of proteins which exist widely in nature having been isolated from a variety of organisms as described e.g., in Nelson, *supra*. These organisms include various mammals, fish, invertebrates, plants, mollusk, crustaceans, lower eukaryotes and bacteria (Nelson, *supra*). First discovered in rodent liver microsomes as a carbon-monoxide binding pigment as described, e.g., in Garfinkel, *Arch. Biochem. Biophys.* 77:493-509 (1958), which is incorporated herein by reference, P450s were later named based on their absorption at 450 nm in a reduced-CO coupled difference spectrum as described, e.g., in Omura et al., *J. Biol. Chem.* 239:2370-2378 (1964), which is incorporated herein by reference.

- P450s catalyze the metabolism of a variety of endogenous and exogenous compounds (Nelson, *supra*). Endogenous compounds include steroids, prostanoids, eicosanoids, fat-soluble vitamins, fatty acids, mammalian alkaloids, leukotrienes, biogenic amines and phytotoxins (Nelson, *supra*). P450 metabolism involves such reactions as epoxidation, hydroxylation, dealkylation, N-hydroxylation, sulfoxidation, desulfuration and reductive dehalogenation. These reactions generally make the compound more water soluble, which is conducive for excretion, and more electrophilic. These electrophilic products can have detrimental effects if they react with DNA or other cellular constituents. However, they can react

through conjugation with low molecular weight hydrophilic substances resulting in glucoronidation, sulfation, acetylation, amino acid conjugation or glutathione conjugation typically leading to inactivation and elimination as described, e.g., in Klaassen et al., *Toxicology*, 3rd ed, Macmillan, New York, 1986, incorporated herein by reference.

- 5 P450s are heme thiolate proteins consisting of a heme moiety bound to a single polypeptide chain of 45,000 to 55,000 Da. The iron of the heme prosthetic group is located at the center of a protoporphyrin ring. Four ligands of the heme iron can be attributed to the porphyrin ring. The fifth ligand is a thiolate anion from a cysteinyl residue of the polypeptide. The sixth ligand is probably a hydroxyl group from an amino acid residue, or a moiety with a similar field strength such as a water molecule as described, e.g., in Goeptar et al., *Critical Reviews in Toxicology* 25(1):25-65 (1995), incorporated herein by reference.

- Monooxygenation reactions catalyzed by cytochromes P450 in a eukaryotic membrane-bound system require the transfer of electrons from NADPH to P450 via NADPH-cytochrome P450 reductase (*CPR*) as described, e.g., in Taniguchi et al., *Arch. Biochem. Biophys.* 232:585 (1984), incorporated herein by reference. *CPR* is a flavoprotein of approximately 78,000 Da containing 1 mol of flavin adenine dinucleotide (FAD) and 1 mol of flavin mononucleotide (FMN) per mole of enzyme as described, e.g., in Potter et al., *J. Biol. Chem.* 258:6906 (1983), incorporated herein by reference. The FAD moiety of *CPR* is the site of electron entry into the enzyme, whereas FMN is the electron-donating site to P450 as described, e.g., in Vermilion et al., *J. Biol. Chem.* 253:8812 (1978), incorporated herein by reference. The overall reaction is as follows:



- 25 Binding of a substrate to the catalytic site of P450 apparently results in a conformational change initiating electron transfer from *CPR* to P450. Subsequent to the transfer of the first electron, O_2 binds to the Fe_2^+ -P450 substrate complex to form Fe_3^+ -P450-substrate complex. This complex is then reduced by a second electron from *CPR*, or, in some cases, NADH via cytochrome b5 and NADH-cytochrome b5 reductase as described, e.g., in Guengerich et al., *Arch. Biochem. Biophys.* 205:365 (1980), incorporated herein by reference. One atom of this reactive oxygen is introduced into the substrate, while the other is reduced to water. The

oxygenated substrate then dissociates, regenerating the oxidized form of the cytochrome P450 as described, e.g., in Klassen, Amdur and Doull, *Casarett and Doull's Toxicology*, Macmillan, New York (1986), incorporated herein by reference.

The P450 reaction cycle can be short-circuited in such a way that O₂ is reduced to O₂⁻ and/or H₂O₂ instead of being utilized for substrate oxygenation. This side reaction is often referred to as the "uncoupling" of cytochrome P450 as described, e.g., in Kuthen et al., *Eur. J. Biochem.* 126:583 (1982) and Poulos et al., *FASEB J.* 6:674 (1992), both of which are incorporated herein by reference. The formation of these oxygen radicals may lead to oxidative cell damage as described, e.g., in Mukhopadhyay, *J. Biol. Chem.* 269(18):13390-13397 (1994) and Ross et al., *Biochem. Pharm.* 49(7):979-989 (1995), both of which are incorporated herein by reference. It has been proposed that cytochrome b5's effect on P450 binding to the CPR results in a more stable complex which is less likely to become "uncoupled" as described, e.g., in Yamazaki et al., *Arch. Biochem. Biophys.* 325(2):174-182 (1996), incorporated herein by reference.

P450 families are assigned based upon protein sequence comparisons. Notwithstanding a certain amount of heterogeneity, a practical classification of P450s into families can be obtained based on deduced amino acid sequence similarity. P450s with amino acid sequence similarity of between about 40 - 80% are considered to be in the same family, with sequences of about > 55% belonging to the same subfamily. Those with sequence similarity of about < 40% are generally listed as members of different P450 gene families (Nelson, *supra*). A value of about > 97% is taken to indicate allelic variants of the same gene, unless proven otherwise based on catalytic activity, sequence divergence in non-translated regions of the gene sequence, or chromosomal mapping.

The most highly conserved region is the HR2 consensus containing the invariant cysteine residue near the carboxyl terminus which is required for heme binding as described, e.g., in Gotoh et al. *J. Biochem.* 93:807-817 (1983) and Motohashi et al., *J. Biochem.* 101:879-997 (1987), both of which are incorporated herein by reference. Additional consensus regions, including the central region of helix I and the transmembrane region, have also been identified, as described, e.g., in Goeptar et al., *supra* and Kalb et al., *PNAS.* 85:7221-7225 (1988), incorporated herein by reference, although the HR2 cysteine is the only invariant amino acid among P450s.

Short chain ($\leq C12$) aliphatic dicarboxylic acids (diacids) are important industrial intermediates in the manufacture of diesters and polymers, and find application as thermoplastics, plasticizing agents, lubricants, hydraulic fluids, agricultural chemicals, pharmaceuticals, dyes, surfactants, and adhesives. The high price and limited availability of short chain diacids are due to constraints imposed by the existing chemical synthesis.

Long-chain diacids (aliphatic α, ω -dicarboxylic acids with carbon numbers of 12 or greater, hereafter also referred to as diacids) ($HOOC-(CH_2)_n-COOH$) are a versatile family of chemicals with demonstrated and potential utility in a variety of chemical products including plastics, adhesives, and fragrances. Unfortunately, the full market potential of diacids has not been realized because chemical processes produce only a limited range of these materials at a relatively high price. In addition, chemical processes for the production of diacids have a number of limitations and disadvantages. All the chemical processes are restricted to the production of diacids of specific carbon chain lengths. For example, the dodecanedioic acid process starts with butadiene. The resulting product diacids are limited to multiples of four-carbon lengths and, in practice, only dodecanedioic acid is made. The dodecanedioic process is based on nonrenewable petrochemical feedstocks. The multireaction conversion process produces unwanted byproducts, which result in yield losses, NO_x pollution and heavy metal wastes.

Long-chain diacids offer potential advantages over shorter chain diacids, but their high selling price and limited commercial availability prevent widespread growth in many of these applications. Biocatalysis offers an innovative way to overcome these limitations with a process that produces a wide range of diacid products from renewable feedstocks. However, there is no commercially viable bioprocess to produce long chain diacids from renewable resources.

SUMMARY OF THE INVENTION

An isolated nucleic acid is provided which encodes a *CPRA* protein having the amino acid sequence set forth in SEQ ID NO: 83. An isolated nucleic acid is also provided which includes a coding region defined by nucleotides 1006-3042 as set forth in SEQ ID NO: 81. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 83. A vector is provided which includes a nucleotide sequence encoding *CPRA* protein

including an amino acid sequence as set forth in SEQ ID NO: 83. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CPRA* protein having an amino acid sequence as set forth in SEQ ID NO: 83. A method of producing a *CPRA* protein including an amino acid sequence as set forth in SEQ ID NO: 83 is also provided which includes a)

- 5 transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 83; and b) culturing the cell under conditions favoring the expression of the protein.

- An isolated nucleic acid is provided which encodes a *CPRB* protein having the amino acid sequence set forth in SEQ ID NO: 84. An isolated nucleic acid is provided which
10 includes a coding region defined by nucleotides 1033-3069 as set forth in SEQ ID NO: 82. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 84. A vector is provided which includes a nucleotide sequence encoding *CPRB* protein including an amino acid sequence as set forth in SEQ ID NO: 84. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CPRB* protein having an amino acid
15 sequence as set forth in SEQ ID NO: 84. A method of producing a *CPRB* protein including an amino acid sequence as set forth in SEQ ID NO: 84 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 84; and b) culturing the cell under conditions favoring the expression of the protein.

- 20 An isolated nucleic acid is provided which encodes a *CYP52A1A* protein having the amino acid sequence set forth in SEQ ID NO: 95. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1177-2748 as set forth in SEQ ID NO: 85. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 95. A vector is provided which includes a nucleotide sequence encoding *CYP52A1A* protein
25 including an amino acid sequence as set forth in SEQ ID NO: 95. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A1A* protein having an amino acid sequence as set forth in SEQ ID NO: 95. A method of producing a *CYP52A1A* protein including an amino acid sequence as set forth in SEQ ID NO: 95 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino
30 acid sequence as set forth in SEQ ID NO: 95; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A2A* protein is provided which has the amino acid sequence set forth in SEQ ID NO: 96. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1199-2767 as set forth in SEQ ID NO: 86. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 96. A vector is provided which includes a nucleotide sequence encoding *CYP52A2A* protein including an amino acid sequence as set forth in SEQ ID NO: 96. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A2A* protein having an amino acid sequence as set forth in SEQ ID NO: 96. A method of producing a *CYP52A2A* protein including an amino acid sequence as set forth in SEQ ID NO: 96 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 96; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A2B* protein is provided which has the amino acid sequence set forth in SEQ ID NO: 97. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1072-2640 as set forth in SEQ ID NO: 87. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 97. A vector is provided which includes a nucleotide sequence encoding *CYP52A2B* protein including an amino acid sequence as set forth in SEQ ID NO: 97. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A2B* protein having an amino acid sequence as set forth in SEQ ID NO: 97. A method of producing a *CYP52A2B* protein including an amino acid sequence as set forth in SEQ ID NO: 97 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 97; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A3A* protein is provided which has the amino acid sequence set forth in SEQ ID NO: 98. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1126-2748 as set forth in SEQ ID NO: 88. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 98. A vector is provided which includes a nucleotide sequence encoding *CYP52A3A* protein including an amino acid sequence as set forth in SEQ ID NO: 98. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A3A* protein having an

amino acid sequence as set forth in SEQ ID NO: 98. A method of producing a *CYP52A3A* protein including an amino acid sequence as set forth in SEQ ID NO: 98 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 98; and b) culturing the cell under
5 conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A3B* protein is provided having the amino acid sequence as set forth in SEQ ID NO: 99. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 913-2535 as set forth in SEQ ID NO: 89. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO:
10 99. A vector is provided which includes a nucleotide sequence encoding *CYP52A3B* protein including an amino acid sequence as set forth in SEQ ID NO: 99. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A3B* protein having an amino acid sequence as set forth in SEQ ID NO: 99. A method of producing a *CYP52A3B* protein including an amino acid sequence as set forth in SEQ ID NO: 99 is provided which includes a)
15 transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 99; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A5A* protein is provided having the amino acid sequence set forth in SEQ ID NO: 100. An isolated nucleic acid is provided which
20 includes a coding region defined by nucleotides 1103-2656 as set forth in SEQ ID NO: 90. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 100. A vector is provided which includes a nucleotide sequence encoding *CYP52A5A* protein including an amino acid sequence as set forth in SEQ ID NO: 100. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A5A* protein having an
25 amino acid sequence as set forth in SEQ ID NO: 100. A method of producing a *CYP52A5A* protein including an amino acid sequence as set forth in SEQ ID NO: 100 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 100; and b) culturing the cell under conditions favoring the expression of the protein.

30 An isolated nucleic acid encoding a *CYP52A5B* protein is provided having the amino acid sequence as set forth in SEQ ID NO: 101. An isolated nucleic acid is provided

which includes a coding region defined by nucleotides 1142-2695 as set forth in SEQ ID NO: 91. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 101. A vector is provided which includes a nucleotide sequence encoding *CYP52A5B* protein including the amino acid sequence as set forth in SEQ ID NO: 101. A host cell is
5 provided which is transfected or transformed with the nucleic acid encoding *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101. A method of producing a *CYP52A5B* protein including an amino acid sequence as set forth in SEQ ID NO: 101 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 101; and b) culturing the
10 cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A8A* protein is provided having the amino acid sequence set forth in SEQ ID NO: 102. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 464-2002 as set forth in SEQ ID NO: 92. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO:
15 102. A vector is provided which includes a nucleotide sequence encoding *CYP52A8A* protein including an amino acid sequence as set forth in SEQ ID NO: 102. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A8A* protein having an amino acid sequence as set forth in SEQ ID NO: 102. A method of producing a *CYP52A8A* protein including an amino acid sequence as set forth in SEQ ID NO: 102 is provided which
20 includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 102; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A8B* protein is provided having the amino acid sequence set forth in SEQ ID NO: 103. An isolated nucleic acid is provided which
25 includes a coding region defined by nucleotides 1017-2555 as set forth in SEQ ID NO: 93. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 103. A vector is provided which includes a nucleotide sequence encoding *CYP52A8B* protein including an amino acid sequence as set forth in SEQ ID NO: 103. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A8B* protein having an
30 amino acid sequence as set forth in SEQ ID NO: 103. A method of producing a *CYP52A8B* protein including an amino acid sequence as set forth in SEQ ID NO: 103 is provided which

includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 103; and b) culturing the cell under conditions favoring the expression of the protein.

5 An isolated nucleic acid encoding a *CYP52D4A* protein is provided having the amino acid sequence set forth in SEQ ID NO: 104. An isolated nucleic acid is provided including a coding region defined by nucleotides 767-2266 as set forth in SEQ ID NO: 94. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 104. A vector is provided which includes a nucleotide sequence encoding *CYP52D4A* protein including an amino acid sequence as set forth in SEQ ID NO: 104. A host cell is provided
10 which is transfected or transformed with the nucleic acid encoding *CYP52D4A* protein having an amino acid sequence as set forth in SEQ ID NO: 104. A method of producing a *CYP52D4A* protein including an amino acid sequence as set forth in SEQ ID NO: 104 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 104; and b) culturing the cell under
15 conditions favoring the expression of the protein.

A method for discriminating members of a gene family by quantifying the amount of target mRNA in a sample is provided which includes a) providing an organism containing a target gene; b) culturing the organism with an organic substrate which causes upregulation in the activity of the target gene; c) obtaining a sample of total RNA from the organism at a first point
20 in time; d) combining at least a portion of the sample of the total RNA with a known amount of competitor RNA to form an RNA mixture, wherein the competitor RNA is substantially similar to the target mRNA but has a lesser number of nucleotides compared to the target mRNA; e) adding reverse transcriptase to the RNA mixture in a quantity sufficient to form corresponding target DNA and competitor DNA; (f) conducting a polymerase chain reaction in the presence of
25 at least one primer specific for at least one substantially non-homologous region of the target DNA within the gene family, the primer also specific for the competitor DNA; g) repeating steps (c-f) using increasing amounts of the competitor RNA while maintaining a substantially constant amount of target RNA; h) determining the point at which the amount of target DNA is substantially equal to the amount of competitor DNA; i) quantifying the results by comparing the
30 ratio of the concentration of unknown target to the known concentration of competitor; and j)

obtaining a sample of total RNA from the organism at another point in time and repeating steps (d-i).

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CPRA* genes; b) increasing, in the host cell, the number of *CPRA* genes which encode a *CPRA* protein having the amino acid sequence as set forth in SEQ ID NO: 83; c) culturing the host cell in media containing an organic substrate which upregulates the *CPRA* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CPRA* protein having an amino acid sequence as set forth in SEQ ID NO: 83 is provided which includes a) transforming a host cell having a naturally occurring amount of *CPRA* protein with an increased copy number of a *CPRA* gene that encodes the *CPRA* protein having the amino acid sequence as set forth in SEQ ID NO: 83; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CPRA* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CPRB* genes; b) increasing, in the host cell, the number of *CPRB* genes which encode a *CPRB* protein having the amino acid sequence as set forth in SEQ ID NO: 84; c) culturing the host cell in media containing an organic substrate which upregulates the *CPRB* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CPRB* protein having an amino acid sequence as set forth in SEQ ID NO: 84 is provided which includes a) transforming a host cell having a naturally occurring amount of *CPRB* protein with an increased copy number of a *CPRB* gene that encodes the *CPRB* protein having the amino acid sequence as set forth in SEQ ID NO: 84; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CPRB* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A1A* genes; b) increasing, in the host cell, the number of *CYP52A1A* genes which encode a *CYP52A1A* protein having the amino acid sequence as set forth in SEQ ID NO: 95; c) culturing the host cell in media

containing an organic substrate which upregulates the *CYP52A1A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A1A* protein having an amino acid sequence as set forth in SEQ ID NO: 95 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A1A* protein with an increased copy number of a *CYP52A1A* gene that encodes the *CYP52A1A* protein having the amino acid sequence as set forth in SEQ ID NO: 95; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A1A* gene.

10 A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A2A* genes; b) increasing, in the host cell, the number of *CYP52A2A* genes which encode a *CYP52A2A* protein having the amino acid sequence as set forth in SEQ ID NO: 96; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A2A* protein having an amino acid sequence as set forth in SEQ ID NO: 96 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A2A* protein with an increased copy number of a *CYP52A2A* gene that encodes the *CYP52A2A* protein having the amino acid sequence as set forth in SEQ ID NO: 96; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A2A* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A2B* genes; b) increasing, in the host cell, the number of *CYP52A2B* genes which encode a *CYP52A2B* protein having the amino acid sequence as set forth in SEQ ID NO: 97; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2B* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A2B* protein having an amino acid sequence as set forth in SEQ ID NO: 97 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A2B* protein with an increased copy number of

a *CYP52A2B* gene that encodes the *CYP52A2B* protein having the amino acid sequence as set forth in SEQ ID NO: 97; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A2B* gene.

5 A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A3A* genes; b) increasing, in the host cell, the number of *CYP52A3A* genes which encode a *CYP52A3A* protein having the amino acid sequence as set forth in SEQ ID NO: 98; c) culturing the host cell in media containing an organic substrate which upregulates *CYP52A3A* gene, to effect increased
10 production of dicarboxylic acid.

 A method for increasing the production of a *CYP52A3A* protein having an amino acid sequence as set forth in SEQ ID NO: 98 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A3A* protein with an increased copy number of a *CYP52A3A* gene that encodes the *CYP52A3A* protein having the amino acid sequence as set
15 forth in SEQ ID NO: 98; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A3A* gene.

 A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A3B* genes; b)
20 increasing, in the host cell, the number of *CYP52A3B* genes which encode a *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A3B* gene, to effect increased production of dicarboxylic acid.

 A method for increasing the production of a *CYP52A3B* protein having an amino acid sequence as set forth in SEQ ID NO: 99 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A3B* protein with an increased copy number of a *CYP52A3B* gene that encodes the *CYP52A3B* protein having the amino acid sequence as set
25 forth in SEQ ID NO: 99; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the
30 *CYP52A3B* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A5A* genes; b) increasing, in the host cell, the number of *CYP52A5A* genes which encode a *CYP52A5A* protein having the amino acid sequence as set forth in SEQ ID NO: 100; c) culturing the host cell in
5 media containing an organic substrate which upregulates the *CYP52A5A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A5A* protein having an amino acid sequence as set forth in SEQ ID NO: 100 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A5A* protein with an increased copy number of
10 a *CYP52A5A* gene that encodes the *CYP52A5A* protein having the amino acid sequence as set forth in SEQ ID NO: 100; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A5A* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A5B* genes; b)
15 increasing, in the host cell, the number of *CYP52A5B* genes which encode a *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A5B* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A5B* protein having an amino acid sequence as set forth in SEQ ID NO: 101 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A5B* protein with an increased copy number of
20 a *CYP52A5B* gene that encodes the *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A5B* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A8A* genes; b)
increasing, in the host cell, the number of *CYP52A8A* genes which encode a *CYP52A8A* protein
30 having the amino acid sequence as set forth in SEQ ID NO: 102; c) culturing the host cell in

media containing an organic substrate which upregulates the *CYP52A8A* gene, to effect increased production of dicarboxylic acid.

5 A method for increasing the production of a *CYP52A8A* protein having an amino acid sequence as set forth in SEQ ID NO: 102 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A8A* protein with an increased copy number of a *CYP52A8A* gene that encodes the *CYP52A8A* protein having the amino acid sequence as set forth in SEQ ID NO: 102; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A8A* gene.

10 A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A8B* genes; b) increasing, in the host cell, the number of *CYP52A8B* genes which encode a *CYP52A8B* protein having the amino acid sequence as set forth in SEQ ID NO: 103; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A8B* gene, to effect increased production of dicarboxylic acid.

15 A method for increasing the production of a *CYP52A8B* protein having an amino acid sequence as set forth in SEQ ID NO: 103 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A8B* protein with an increased copy number of a *CYP52A8B* gene that encodes the *CYP52A8B* protein having the amino acid sequence as set forth in SEQ ID NO: 103; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A8B* gene.

20 A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52D4A* genes; b) increasing, in the host cell, the number of *CYP52D4A* genes which encode a *CYP52D4A* protein having the amino acid sequence as set forth in SEQ ID NO: 104; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52D4A* gene, to effect increased production of dicarboxylic acid.

25 A method for increasing the production of a *CYP52D4A* protein having an amino acid sequence as set forth in SEQ ID NO: 104 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52D4A* protein with an increased copy number

of a *CYP52D4A* gene that encodes the *CYP52D4A* protein having the amino acid sequence as set forth in SEQ ID NO: 104; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52D4A* gene.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation of cloning vector pTriplEx from Clontech™ Laboratories, Inc. Selected restriction sites within the multiple cloning site are shown.

10

Figure 2A is a map of the ZAP Express™ vector.

Figure 2B is a schematic representation of cloning phagemid vector pBK-CMV.

Figure 3 is a double stranded DNA sequence of a portion of the 5 prime coding region of the *CYP52A5A* gene (SEQ ID NO: 36).

Figure 4 is a diagrammatic representation of highly conserved regions of *CYP* and *CPR* gene protein sequences. Helix I represents the putative substrate binding site and HR2 represents the heme binding region. The FMN, FAD and NADPH binding regions are indicated below the *CPR* gene.

15

Figure 5 is a diagrammatic representation of the plasmid pHKM1 containing the truncated *CPRA* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

20

Figure 6 is a diagrammatic representation of the plasmid pHKM4 containing the truncated *CPRA* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

25

Figure 7 is a diagrammatic representation of the plasmid pHKM9 containing the *CPRB* gene (SEQ ID NO: 82) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

30

Figure 8 is a diagrammatic representation of the plasmid pHKM11 containing the *CYP52A1A* gene (SEQ ID NO: 85) present in the pBK-CMV vector. A detailed restriction map

of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 9 is a diagrammatic representation of the plasmid pHKM12 containing the *CYP52A8A* gene (SEQ ID NO: 92) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 10 is a diagrammatic representation of the plasmid pHKM13 containing the *CYP52D4A* gene (SEQ ID NO: 94) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 11 is a diagrammatic representation of the plasmid pHKM14 containing the *CYP52A2B* gene (SEQ ID NO: 87) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 12 is a diagrammatic representation of the plasmid pHKM15 containing the *CYP52A8B* gene (SEQ ID NO: 93) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figures 13A-13D show the complete DNA sequences including regulatory and coding regions for the *CPRA* gene (SEQ ID NO: 81) and *CPRB* gene (SEQ ID NO: 82) from *C. tropicalis* ATCC 20336. Figures 13A-13D show regulatory and coding region alignment of these sequences. Asterisks indicate conserved nucleotides. Bold indicates protein coding nucleotides; the start and stop codons are underlined.

Figure 14 shows the amino acid sequence of the *CPRA* (SEQ ID NO: 83) and *CPRB* (SEQ ID NO: 84) proteins from *C. tropicalis* ATCC 20336 and alignment of these amino acid sequences. Asterisks indicate residues which are not conserved.

Figures 15A-15M show the complete DNA sequences including regulatory and coding regions for the following genes from *C. tropicalis* ATCC 20366: *CYP52A1A* (SEQ ID NO: 85), *CYP52A2A* (SEQ ID NO: 86), *CYP52A2B* (SEQ ID NO: 87), *CYP52A3A* (SEQ ID NO: 88), *CYP52A3B* (SEQ ID NO: 89), *CYP52A5A* (SEQ ID NO: 90), *CYP52A5B* (SEQ ID NO: 91), *CYP52A8A* (SEQ ID NO: 92), *CYP52A8B* (SEQ ID NO: 93), and *CYP52D4A* (SEQ ID NO: 94).

Figures 15A-15M show regulatory and coding region alignment of these sequences. Asterisks indicate conserved nucleotides. Bold indicates protein coding nucleotides; the start and stop codons are underlined.

Figures 16A-16C show the amino acid sequences encoding the *CYP52A1A* (SEQ ID NO: 95), *CYP52A2A* (SEQ ID NO: 96), *CYP52A2B* (SEQ ID NO: 97), *CYP52A3A* (SEQ ID NO: 98), *CYP52A3B* (SEQ ID NO: 99), *CYP52A5A* (SEQ ID NO: 100), *CYP52A5B* (SEQ ID NO: 101), *CYP52A8A* (SEQ ID NO: 102), *CYP52A8B* (SEQ ID NO: 103) and *CYP52D4A* (SEQ ID NO: 104) proteins from *C. tropicalis* ATCC 20336. Asterisks indicate identical residues and dots indicate conserved residues.

Figure 17 is a diagrammatic representation of the pTag PCR product cloning vector (commercially available from R&D Systems, Minneapolis, MN).

Figure 18 is a plot of the log ratio (U/C) of unknown target DNA product to competitor DNA product versus the concentration of competitor mRNA. The plot is used to calculate the target messenger RNA concentration in a quantitative competitive reverse transcription polymerase chain reaction (QC-RT-PCR).

Figure 19 is a graph showing the relative induction of *C. tropicalis* ATCC 20962 *CYP52A5A* (SEQ ID NO: 90) by the addition of the fatty acid substrate Emersol® 267 to the growth medium.

Figure 20 is a graph showing the induction of *C. tropicalis* ATCC 20962 *CYP52* and *CPR* genes by Emersol® 267. P450 genes *CYP52A3A* (SEQ ID NO: 88), *CYP52A3B* (SEQ ID NO: 89), and *CYP52D4A* (SEQ ID NO: 94) are expressed at levels below the detection level of the QC-RT-PCR assay.

Figure 21 is a scheme to integrate selected genes into the genome of *Candida tropicalis* strains and recovery of *URA3A* selectable marker.

Figure 22 is a schematic representation of the transformation of *C. tropicalis* H5343 *ura3⁻* with *CYP* and/or *CPR* genes. Only one *URA3* locus needs to be functional. There are a total of 6 possible *ura3* targets (*Sura3A* loci-2 *pox4* disruptions, 2 *pox 5* disruptions, 1 *ura3A* locus; and 1 *ura3B* locus).

Figure 23 is the complete DNA sequence (SEQ ID NO: 105) encoding *URA3A* from *C. tropicalis* ATCC 20336 and the amino acid sequence of the encoded protein (SEQ ID NO: 106).

Figure 24 is a schematic representation of the plasmid pURain, the base vector for integrating selected genes into the genome of *C. tropicalis*. The detailed construction of pURain is described in the text.

Figure 25 is a schematic representation of the plasmid pNEB193 cloning vector
5 (commercially available from New England Biolabs, Beverly, MA).

Figure 26 is a diagrammatic representation of the plasmid pPA15 containing the truncated *CYP52A2A* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

10 Figure 27 is a schematic representation of pURA2in, the base vector is constructed in pNEB193 which contains the 8 bp recognition sequences for *Asc I*, *Pac I* and *Pme I*. *URA3A* (SEQ ID NO: 105) and *CYP52A2A* (SEQ ID NO: 86) do not contain these 8 bp recognition sites. *URA3A* is inverted so that the transforming fragment will attempt to recircularize prior to integration. An *Asc I/Pme I* fragment was used to transform H5343 *ura*⁻.

15 Figure 28 shows a scheme to detect integration of *CYP52A2A* gene (SEQ ID NO: 86) into the genome of H5343 *ura*⁻. In all cases, hybridization band intensity could reflect the number of integrations.

Figure 29 is a diagrammatic representation of the plasmid pPA57 containing the truncated *CYP52A3A* gene present in the pTriplEx vector. A detailed restriction map of only the
20 sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 30 is a diagrammatic representation of the plasmid pPA62 containing the truncated *CYP52A3B* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of
25 transcription is indicated by an arrow under the open reading frame.

Figure 31 is a diagrammatic representation of the plasmid pPAL3 containing the truncated *CYP52A5A* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

30 Figure 32 is a diagrammatic representation of the plasmid pPA5 containing the truncated *CYP52A5A* gene present in the pTriplEx vector. A detailed restriction map of only the

sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 33 is a diagrammatic representation of the plasmid pPA18 containing the truncated *CYP52D4A* gene present in the pTriplEx vector. A detailed restriction map of only the
5 sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 34 is a graph showing the expression of *CYP52A1* (SEQ ID NO: 85), *CYP52A2* (SEQ ID NO: 86) and *CYP52A5* genes (SEQ ID NOS: 90 and 91) from *C. tropicalis* 20962 in a fermentor run upon the addition of amounts of the substrate oleic acid or tridecane in
10 a spiking experiment.

Figure 35 depicts a scheme used for the extraction and analysis of diacids and monoacids from fermentation broths.

Figure 36 is a graph showing the induction of expression of *CYP52A1A*, *CYP52A2A* and *CYP52A5A* in a fermentor run upon addition of the substrate octadecane. No
15 induction of *CYP52A3A* or *CYP52A3B* was observed under these conditions.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Diacid productivity is improved according to the present invention by selectively increasing enzymes which are known to be important to the oxidation of organic substrates such
20 as fatty acids composing the desired feed. According to the present invention, ten *CYP* genes and two *CPR* genes of *C. tropicalis* have been identified and characterized that relate to participation in the ω -hydroxylase complex catalyzing the first step in the ω -oxidation pathway. In addition, a novel quantitative competitive reverse transcription polymerase chain reaction (QC-RT-PCR) assay is used to measure gene expression in the fermentor under conditions of
25 induction by one or more organic substrates as defined herein. Based upon QC-RT-PCR results, three *CYP* genes, *CYP52A1*, *CYP52A2* and *CYP52A5*, have been identified as being of greater importance for the ω -oxidation of long chain fatty acids. Amplification of the *CPR* gene copy number improves productivity. The QC-RT-PCR assay indicates that both *CYP* and *CPR* genes appear to be under tight regulatory control.

30 In accordance with the present invention, a method for discriminating members of a gene family by quantifying the amount of target mRNA in a sample is provided which

includes a) providing an organism containing a target gene; b) culturing the organism with an organic substrate which causes upregulation in the activity of the target gene; c) obtaining a sample of total RNA from the organism at a first point in time; d) combining at least a portion of the sample of the total RNA with a known amount of competitor RNA to form an RNA mixture, wherein the competitor RNA is substantially similar to the target mRNA but has a lesser number of nucleotides compared to the target mRNA; e) adding reverse transcriptase to the RNA mixture in a quantity sufficient to form corresponding target DNA and competitor DNA; (f) conducting a polymerase chain reaction in the presence of at least one primer specific for at least one substantially non-homologous region of the target DNA within the gene family, the primer also specific for the competitor DNA; g) repeating steps (c-f) using increasing amounts of the competitor RNA while maintaining a substantially constant amount of target RNA; h) determining the point at which the amount of target DNA is substantially equal to the amount of competitor DNA; i) quantifying the results by comparing the ratio of the concentration of unknown target to the known concentration of competitor; and j) obtaining a sample of total RNA from the organism at another point in time and repeating steps (d-i).

In addition, modification of existing promoters and/or the isolation of alternative promoters provides increased expression of *CYP* and *CPR* genes. Strong promoters are obtained from at least four sources: random or specific modifications of the *CYP52A2* promoter, *CYP52A5* promoter, *CYP52A1* promoter, the selection of a strong promoter from available *Candida* β -oxidation genes such as *POX4* and *POX5*, or screening to select another suitable *Candida* promoter.

Promoter strength can be directly measured using QT-RT-PCR to measure *CYP* and *CPR* gene expression in *Candida* cells isolated from fermentors. Enzymatic assays and antibodies specific for *CYP* and *CPR* proteins are used to verify that increased promoter strength is reflected by increased synthesis of the corresponding enzymes. Once a suitable promoter is identified, it is fused to the selected *CYP* and *CPR* genes and introduced into *Candida* for construction of a new improved production strain. It is contemplated that the coding region of the *CYP* and *CPR* genes can be fused to suitable promoters or other regulatory sequences which are well known to those skilled in the art.

In accordance with the present invention, studies on *C. tropicalis* ATCC 20336 have identified six unique *CYP* genes and four potential alleles. QC-RT-PCR analyses of cells

isolated during the course of the fermentation bioconversions indicate that at least three of the *CYP* genes are induced by fatty acids and at least two of the *CYP* genes are induced by alkanes. See Figure 34. Two of the *CYP* genes are highly induced indicating participation in the ω -hydroxylase complex which catalyzes the rate limiting step in the oxidation of fatty acids to the
5 corresponding diacids.

The biochemical characterizations of each P450 enzyme herein is used to tailor the *C. tropicalis* host for optimal diacid productivity and is used to select P450 enzymes to be amplified based upon the fatty acid content of the feedstream. *CYP* gene(s) encoding P450 enzymes that have a low specific activity for the fatty acid or alkane substrate of choice are
10 targeted for inactivation, thereby reducing the physiological load on the cell.

Since it has been demonstrated that *CPR* can be limiting in yeast systems, the removal of non-essential P450s from the system can free electrons that are being used by non-essential P450s and make them available to the P450s important for diacid productivity. Moreover, the removal of non-essential P450s can make available other necessary but potentially
15 limiting components of the P450 system (i.e., available membrane space, heme and/or NADPH).

Diacid productivity is thus improved by selective integration, amplification, and over expression of *CYP* and *CPR* genes in the *C. tropicalis* production host.

It should be understood that host cells into which one or more copies of desired *CYP* and/or *CPR* genes have been introduced can be made to include such genes by any
20 technique known to those skilled in the art. For example, suitable host cells include procaryotes such as *Bacillus* sp., *Pseudomonus* sp., *Actinomycetes* sp., *Eschericia* sp., *Mycobacterium* sp., and eukaryotes such as yeast, algae, insect cells, plant cells and and filamentous fungi. Suitable host cells are preferably yeast cells such as *Yarrowia*, *Bebaromyces*, *Saccharomyces*, *Schizosaccharomyces*, and *Pichia* and more preferably those of the *Candida* genus. Preferred
25 species of *Candida* are *tropicalis*, *maltosa*, *apicola*, *paratropicalis*, *albicans*, *cloacae*, *guilliermondii*, *intermedia*, *lipolytica*, *parapsilosis* and *zeylenoides*. Certain preferred stains of *Candida tropicalis* are listed in U.S. Patent No. 5,254,466, incorporated herein by reference.

Vectors such as plasmids, phagemids, phages or cosmids can be used to transform or transfect suitable host cells. Host cells may also be transformed by introducing into a cell a
30 linear DNA vector(s) containing the desired gene sequence. Such linear DNA may be advantageous when it is desirable to avoid introduction of non-native (foreign) DNA into the

cell. For example, DNA consisting of a desired target gene(s) flanked by DNA sequences which are native to the cell can be introduced into the cell by electroporation, lithium acetate transformation, spheroplasting and the like. Flanking DNA sequences can include selectable markers and/or other tools for genetic engineering.

5 A suitable organic substrate herein can be any organic compound that is biooxidizable to a mono- or polycarboxylic acid. Such a compound can be any saturated or unsaturated aliphatic compound or any carbocyclic or heterocyclic aromatic compound having at least one terminal methyl group, a terminal carboxyl group and/or a terminal functional group which is oxidizable to a carboxyl group by biooxidation. A terminal functional group which is a
10 derivative of a carboxyl group may be present in the substrate molecule and may be converted to a carboxyl group by a reaction other than biooxidation. For example, if the terminal group is an ester that neither the wild-type *C. tropicalis* nor the genetic modifications described herein will allow hydrolysis of the ester functionality to a carboxyl group, then a lipase can be added during the fermentation step to liberate free fatty acids. Suitable organic substrates include, but are not
15 limited to, saturated fatty acids, unsaturated fatty acids, alkanes, alkenes, alkynes and combinations thereof.

Alkanes are a type of saturated organic substrate which are useful herein. The alkanes can be linear or cyclic, branched or straight chain, substituted or unsubstituted. Particularly preferred alkanes are those having from about 4 to about 25 carbon atoms, examples
20 of which include but are not limited to butane, hexane, octane, nonane, dodecane, tridecane, tetradecane, octadecane and the like.

Examples of unsaturated organic substrates which can be used herein include but are not limited to internal olefins such as 2-pentene, 2-hexene, 3-hexene, 9-octadecene and the like; unsaturated carboxylic acids such as 2-hexenoic acid and esters thereof, oleic acid and esters
25 thereof including triglyceryl esters having a relatively high oleic acid content, erucic acid and esters thereof including triglyceryl esters having a relatively high erucic acid content, ricinoleic acid and esters thereof including triglyceryl esters having a relatively high ricinoleic acid content, linoleic acid and esters thereof including triglyceryl esters having a relatively high linoleic acid content; unsaturated alcohols such as 3-hexen-1-ol, 9-octadecen-1-ol and the like; unsaturated
30 aldehydes such as 3-hexen-1-al, 9-octadecen-1-al and the like. In addition to the above, an organic substrate which can be used herein include alicyclic compounds having at least one

internal carbon-carbon double bond and at least one terminal methyl group, a terminal carboxyl group and/or a terminal functional group which is oxidizable to a carboxyl group by biooxidation. Examples of such compounds include but are not limited to 3,6-dimethyl, 1,4-cyclohexadiene; 3-methylcyclohexene; 3-methyl-1, 4-cyclohexadiene and the like.

5 Examples of the aromatic compounds that can be used herein include but are not limited to arenes such as o-, m-, p-xylene; o-, m-, p-methyl benzoic acid; dimethyl pyridine, and the like. The organic substrate can also contain other functional groups that are biooxidizable to carboxyl groups such as an aldehyde or alcohol group. The organic substrate can also contain other functional groups that are not biooxidizable to carboxyl groups and do not interfere with
10 the biooxidation such as halogens, ethers, and the like.

 Examples of saturated fatty acids which may be applied to cells incorporating the present *CYP* and *CPR* genes include caproic, enanthic, caprylic, pelargonic, capric, undecylic, lauric, myristic, pentadecanoic, palmitic, margaric, stearic, arachidic, behenic acids and combinations thereof. Examples of unsaturated fatty acids which may be applied to cells
15 incorporating the present *CYP* and *CPR* genes include palmitoleic, oleic, erucic, linoleic, linolenic acids and combinations thereof. Alkanes and fractions of alkanes may be applied which include chain links from C12 to C24 in any combination. An example of a preferred fatty acid mixtures are Emersol® 267 and Tallow, both commercially available from Henkel Chemicals Group, Cincinnati, OH. The typical fatty acid composition of Emersol® 267 and
20 Tallow is as follows:

		<u>TALLOW</u>	<u>E267</u>
25	C14:0	3.5%	2.4%
	C14:1	1.0%	0.7%
	C15:0	0.5%	-----
	C16:0	25.5%	4.6%
	C16:1	4.0%	5.7%
	C17:0	2.5%	-----
	C17:1	-----	5.7%
30	C18:0	19.5%	1.0%
	C18:1	41.0%	69.9%
	C18:2	2.5%	8.8%

C18:3	-----	0.3%
C20:0	0.5%	-----
C20:1	-----	0.9%

5 The following examples are meant to illustrate but not to limit the invention. All relevant microbial strains and plasmids are described in Table 1 and Table 2, respectively.

Table 1. List of *Escherichia coli* and *Candida tropicalis* strains

<i>E. coli</i> STRAIN	GENOTYPE	SOURCE
10 XL1Blue-MRF'	<i>endA1, gyrA96, hsdR17, lac⁻, recA1, relA1, supE44, thi-1, [F' lacFZ M15, proAB, Tn10]</i>	Stratagene, La Jolla, CA
BM25.8	<i>SupE44, thi (lac-proAB) [F' traD36, proAB⁺, lacFZ M15] λimm434 (kan^R)P1 (cam^R) hsdR (r_{h17} m_{h17})</i>	Clontech, Palo Alto, CA
XL0LR	<i>(mcrA)183 (mcrCB-hsdSMR-mrr)173 endA1 thi-1 recA1 gyrA96 relA1 lac [F' proAB lacFZ M15 Tn10 (Tet^r) Su⁻ (nonsuppressing λ'(lambda resistant))</i>	Stratagene, La Jolla, CA

<i>C. tropicalis</i> STRAIN	GENOTYPE	SOURCE
15 ATCC20336	Wild-type	American Type Culture Collection, Rockville, MD
ATCC750	Wild-type	American Type Culture Collection, Rockville, MD
20 ATCC 20962	<i>ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A</i>	Henkel
H5343 ura-	<i>ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3-</i>	Henkel
HDC1	<i>ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CYP52A2A</i>	Henkel
HDC5	<i>ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CYP52A3A</i>	Henkel
25 HDC10	<i>ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CPRB</i>	Henkel

HDC15	<i>ura3A/ura3B</i> , <i>pox4A::ura3A/pox4B::ura3A</i> , <i>pox5::ura3A/pox5::URA3A</i> , <i>ura3::URA3A-CYP52A5A</i>	Henkel
HDC20	<i>ura3A/ura3B</i> , <i>pox4A::ura3A/pox4B::ura3A</i> , <i>pox5::ura3A/pox5::URA3A</i> , <i>ura3::URA3A-CYP52A2A + CPR B</i> (CYP and CPR have opposite 5' to 3' orientation with respect to each other)	Henkel
HDC23	<i>ura3A/ura3B</i> , <i>pox4A::ura3A/pox4B::ura3A</i> , <i>pox5::ura3A/pox5::URA3A</i> , <i>ura3::URA3A-CYP52A2A + CPR B</i> (CYP and CPR have same 5' to 3' orientation with respect to each other)	Henkel

5

Table 2. List of plasmids isolated from genomic libraries and constructed for use in gene integrations.

10

Plasmid	Base vector	Insert	Insert Size	Plasmid size	Description
pURAIin	pNEB193	<i>URA3A</i>	1706 bp	4399 bp	pNEB193 with the <i>URA3A</i> gene inserted in the <i>AscI</i> - <i>PmeI</i> site, generating a <i>PacI</i> site
pURA 2in	pURAIin	<i>CYP52A2A</i>	2230 bp	6629 bp	pURAIin containing a PCR <i>CYP52A2A</i> allele containing <i>PacI</i> restriction sites
pURA REDB in	pURAIin	<i>CPRB</i>	3266 bp	7665 bp	pURAIin containing a PCR <i>CPRB</i> allele containing <i>PacI</i> restriction sites
pHKM1	pTriplEx	Truncated <i>CPRA</i> gene	Approx. 3.8 kb	Approx. 7.4 kb	A truncated <i>CPRA</i> gene obtained by first screening library containing the 5' untranslated region and 1.2 kb open reading frame
pHKM4	PTriplEx	Truncated <i>CPRA</i> gene	Approx. 5 kb	Approx. 8.6 kb	A truncated <i>CPRA</i> gene obtained by screening second library containing the 3' untranslated region end sequence
pHKM9	pBC-CMV	<i>CPRB</i> gene	Approx. 5.3 kb	Approx. 9.8 kb	<i>CPRB</i> allele isolated from the third library
pHKM11	pBC-CMV	<i>CYP52A1A</i>	Approx. 5 kb	Approx. 9.5 kb	<i>CYP52A1A</i> isolated from the third library
pHKM12	pBC-CMV	<i>CYP52A8A</i>	Approx. 7.5 kb	Approx. 12 kb	<i>CYP52A8A</i> isolated from the third library
pHKM13	pBC-CMV	<i>CYP52D4A</i>	Approx. 7.3 kb	Approx. 11.8 kb	<i>CYP52D4A</i> isolated from the third library

15

pHKM14	pBC-CMV	CYP52A2B	Approx. 6 kb	Approx. 10.5 kb	CYP52A2B isolated from the third library
pHKM15	pBC-CMV	CYP52A8B	Approx. 6.6 kb	Approx. 11.1 kb	CYP52A8B isolated from the third library
pPAL3	pTriplEx	CYP52A5A	4.4 kb	Approx. 8.1 kb	CYP52A5A isolated from the 1st library
pPA5	pTriplEx	CYP52A5B	4.1 kb	Approx. 7.8 kb	CYP52A5B isolated from the 2nd library
pPA15	pTriplEx	CYP52A2A	6.0 kb	Approx. 9.7 kb	CYP52A2A isolated from the 2nd library
pPA57	pTriplEx	CYP52A3A	5.5 kb	Approx. 9.2 kb	CYP52A3A isolated from the 2nd library
pPA62	pTriplEx	CYP52A3B	6.0 kb	Approx. 9.7 kb	CYP52A3B isolated from the 2nd library

EXAMPLE 1

10 Purification of Genomic DNA from *Candida tropicalis* ATCC 20336

A. Construction of Genomic Libraries

50 ml of YEPD broth (see Chart) was inoculated with a single colony of *C. tropicalis* 20336 from YEPD agar plate and grown overnight at 30°C. 5 ml of the overnight culture was inoculated into 100 ml of fresh YEPD broth and incubated at 30°C for 4 to 5 hr with shaking. Cells were harvested by centrifugation, washed twice with sterile distilled water and resuspended in 4 ml of spheroplasting buffer (1 M Sorbitol, 50 mM EDTA, 14 mM mercaptoethanol) and incubated for 30 min at 37°C with gentle shaking. 0.5 ml of 2 mg/ml zymolyase (ICN Pharmaceuticals, Inc., Irvine, CA) was added and incubated at 37°C with gentle shaking for 30 to 60 min. Spheroplast formation was monitored by SDS lysis. Spheroplasts were harvested by brief centrifugation (4,000 rpm, 3 min) and were washed once with the spheroplast buffer without mercaptoethanol. Harvested spheroplasts were then suspended in 4 ml of lysis buffer (0.2 M Tris/pH 8.0, 50 mM EDTA, 1% SDS) containing 100 µg/ml RNase (Qiagen Inc., Chatsworth, CA) and incubated at 37°C for 30 to 60 min.

Proteins were denatured and extracted twice with an equal volume of chloroform/isoamyl alcohol (24:1) by gently mixing the two phases by hand inversions. The two phases were separated by centrifugation at 10,000 rpm for 10 min and the aqueous phase containing the high-molecular weight DNA was recovered. To the aqueous layer NaCl was added to a final concentration of 0.2 M and the DNA was precipitated by adding 2 vol of ethanol. Precipitated DNA was spooled with a clean glass rod and resuspended in TE buffer (10 mM

Tris/pH 8.0, 1 mM EDTA) and allowed to dissolve overnight at 4°C. To the dissolved DNA, RNase free of any DNase activity (Qiagen Inc., Chatsworth, CA) was added to a final concentration of 50 µg/ml and incubated at 37°C for 30 min. Then protease (Qiagen Inc., Chatsworth, CA) was added to a final concentration of 100 µg/ml and incubated at 55 to 60°C
5 for 30 min. The solution was extracted once with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) and once with equal volume of chloroform/isoamyl alcohol (24:1). To the aqueous phase 0.1 vol of 3 M sodium acetate and 2 volumes of ice cold ethanol (200 proof) were added and the high molecular weight DNA was spooled with a glass rod and dissolved in 1 to 2 ml of TE buffer.

10

**B. Genomic DNA Preparation for PCR
Amplification of *CYP* and *CPR* Genes**

Five 5 ml of YPD medium was inoculated with a single colony and grown at 30°C overnight. The culture was centrifuged for 5 min at 1200 x g. The supernatant was
15 removed by aspiration and 0.5 ml of a sorbitol solution (0.9 M sorbitol, 0.1 M Tris-Cl pH 8.0, 0.1 M EDTA) was added to the pellet. The pellet was resuspended by vortexing and 1 µl of 2-mercaptoethanol and 50 µl of a 10 µg/ml zymolyase solution were added to the mixture. The tube was incubated at 37°C for 1 hr on a rotary shaker (200 rpm). The tube was then centrifuged for 5 min at 1200 x g and the supernatant was removed by aspiration. The protoplast
20 pellet was resuspended in 0.5 ml 1x TE (10 mM Tris-Cl pH 8.0, 1 mM EDTA) and transferred to a 1.5 ml microcentrifuge tube. The protoplasts were lysed by the addition of 50 µl 10% SDS followed by incubation at 65°C for 20 min. Next, 200 µl of 5M potassium acetate was added and after mixing, the tube was incubated on ice for at least 30 min. Cellular debris was removed by centrifugation at 13,000 x g for 5 min. The supernatant was carefully removed and
25 transferred to a new microfuge tube. The DNA was precipitated by the addition of 1 ml 100% (200 proof) ethanol followed by centrifugation for 5 min at 13,000 x g. The DNA pellet was washed with 1 ml 70 % ethanol followed by centrifugation for 5 min at 13,000 x g. After partially drying the DNA under a vacuum, it was resuspended in 200 µl of 1x TE. The DNA concentration was determined by ratio of the absorbance at 260 nm / 280 nm ($A_{260/280}$).

30

EXAMPLE 2

Construction of *Candida tropicalis* 20336 Genomic Libraries

Three genomic libraries of *C. tropicalis* were constructed, two at Clontech Laboratories, Inc., (Palo Alto, CA) and one at Henkel Corporation (Cincinnati, OH).

5

A. Clontech Libraries

The first Clontech library was made as follows: Genomic DNA was prepared from *C. tropicalis* 20336 as described above, partially digested with *EcoRI* and size fractionated by gel electrophoresis to eliminate fragments smaller than 0.6 kb. Following size fractionation, several ligations of the *EcoRI* genomic DNA fragments and lambda (λ) TriplEx™ vector (Figure 1) arms with *EcoRI* sticky ends were packaged into λ phage heads under conditions designed to obtain one million independent clones. The second genomic library was constructed as follows: Genomic DNA was digested partially with *Sau3A1* and size fractionated by gel electrophoresis. The DNA fragments were blunt ended using standard protocols as described, e.g., in Sambrook et al, *Molecular Cloning: A Laboratory Manual*, 2ed. Cold Spring Harbor Press, USA (1989), incorporated herein by reference. The strategy was to fill in the *Sau3A1* overhangs with Klenow polymerase (Life Technologies, Grand Island, NY) followed by digestion with S1 nuclease (Life Technologies, Grand Island, NY). After S1 nuclease digestion the fragments were end filled one more time with Klenow polymerase to obtain the final blunt-ended DNA fragments. *EcoRI* linkers were ligated to these blunt-ended DNA fragments followed by ligation into the λ TriplEx vector. The resultant library contained approximately 2×10^6 independent clones with an average insert size of 4.5 kb.

10

15

B. Henkel Library

The third genomic library was constructed at Henkel Corporation using λ ZAP Express™ vector (Stratagene, La Jolla, CA) (Figure 2). Genomic DNA was partially digested with *Sau3A1* and fragments in the range of 6 to 12 kb were purified from an agarose gel after electrophoresis of the digested DNA. These DNA fragments were then ligated to *BamHI* digested λ ZAP Express™ vector arms according to manufacturers protocols. Three ligations were set up to obtain approximately 9.8×10^5 independent clones. All three libraries were pooled and amplified according to manufacturer instructions to obtain high-titre ($>10^9$ plaque

20

25

forming units/ml) stock for long-term storage. The titre of packaged phage library was ascertained after infection of *E. coli* XL1Blue-MRF' cells. *E. coli* XL1Blue-MRF' were grown overnight in either in LB medium or NZCYM (Chart) containing 10 mM MgSO₄ and 0.2% maltose at 37°C or 30°C, respectively with shaking. Cells were then centrifuged and resuspended in 0.5 to 1 volume of 10 mM MgSO₄. 200 µl of this *E. coli* culture was mixed with several dilutions of packaged phage library and incubated at 37°C for 15 min. To this mixture 2.5 ml of LB top agarose or NZCYM top agarose (maintained at 60°C) (see Chart) was added and plated on LB agar or NCZYM agar (see Chart) present in 82 mm petri dishes. Phage were allowed to propagate overnight at 37°C to obtain discrete plaques and the phage titre was determined.

EXAMPLE 3

Screening of Genomic Libraries

Both λTriplEx™ and λZAP Express™ vectors are phagemid vectors that can be propagated either as phage or plasmid DNA (after conversion of phage to plasmid). Therefore, the genomic libraries constructed in these vectors can be screened either by plaque hybridization (screening of lambda form of library) or by colony hybridization (screening plasmid form of library after phage to plasmid conversion). Both vectors are capable of expressing the cloned genes and the main difference is the mechanism of excision of plasmid from the phage DNA. The cloning site in λTriplEx™ is located within a plasmid which is present in the phage and is flanked by *loxP* site (Figure 1). When λTriplEx™ is introduced into *E. coli* strain BM25.8 (supplied by Clontech), the *Cre* recombinase present in BM25.8 promotes the excision and circularization of plasmid pTriplEx from the phage λTriplEx™ at the *loxP* sites. The mechanism of excision of plasmid pBK-CMV from phage λZAP Express™ is different. It requires the assistance of a helper phage such as ExAssist™ (Stratagene) and an *E. coli* strain such as XLOR (Stratagene). Both pTriplEx and pBK-CMV can replicate autonomously in *E. coli*.

A. Screening Genomic Libraries (Plasmid Form)

1) Colony Lifts

- A single colony of *E. coli* BM25.8 was inoculated into 5 ml of LB containing 50 µg/ml kanamycin, 10 mM MgSO₄ and 0.1% maltose and grown overnight at 31 °C, 250 rpm. To
- 5 200 µl of this overnight culture (~ 4 X 10⁸ cells) 1 µl of phage library (2 - 5 X 10⁶ plaque forming units) and 150 µl LB broth were added and incubated at 31 °C for 30 min after which 400 µl of LB broth was added and incubated at 31 °C, 225 rpm for 1 h. This bacterial culture was diluted and plated on LB agar containing 50 µg/ml ampicillin (Sigma Chemical Company, St. Louis, MO) and kanamycin (Sigma Chemical Company) to obtain 500 to 600 colonies/plate.
- 10 The plates were incubated at 37 °C for 6 to 7 hrs until the colonies became visible. The plates were then stored at 4 °C for 1.5 h before placing a Colony/Plaque Screen™ Hybridization Transfer Membrane disc (DuPont NEN Research Products, Boston, MA) on the plate in contact with bacterial colonies. The transfer of colonies to the membrane was allowed to proceed for 3 to 5 min. The membrane was then lifted and placed on a fresh LB agar (see Chart) plate containing
- 15 200 µg/ml of chloramphenicol with the side exposed to the bacterial colonies facing up. The plates containing the membranes were then incubated at 37 °C overnight in order to allow full development of the bacterial colonies. The LB agar plates from which colonies were initially lifted were incubated at 37 °C overnight and stored at 4 °C for future use. The following morning the membranes containing bacterial colonies were lifted and placed on two sheets of
- 20 Whatman 3M (Whatman, Hillsboro, OR) paper saturated with 0.5 N NaOH and left at room temperature (RT) for 3 to 6 min to lyse the cells. Additional treatment of membranes was as described in the protocol provided by NEN Research Products.

2) DNA Hybridizations

- 25 Membranes were dried overnight before hybridizing to oligonucleotide probes prepared using a non-radioactive ECL™ 3'-oligolabelling and detection system from Amersham Life Sciences (Arlington Heights, IL). DNA labeling, prehybridization and hybridizations were performed according to manufacturer's protocols. After hybridization, membranes were washed twice at room temperature in 5 X SSC, 0.1% SDS (in a volume equivalent to 2 ml/cm² of
- 30 membrane) for 5 min each followed by two washes at 50 °C in 1X SSC, 0.1% SDS (in a volume

equivalent to 2 ml/cm² of membrane) for 15 min each. The hybridization signal was then generated and detected with Hyperfilm ECLTM (Amersham) according to manufacturer's protocols. Membranes were aligned to plates containing bacterial colonies from which colony lifts were performed and colonies corresponding to positive signals on X-ray were then isolated and propagated in LB broth. Plasmid DNA's were isolated from these cultures and analyzed by restriction enzyme digestions and by DNA sequencing.

B. Screening Genomic Libraries (Plaque Form)

1) λ Library Plating

E. coli XL1Blue-MRF' cells were grown overnight in LB medium (25 ml) containing 10 mM MgSO₄ and 0.2% maltose at 37°C, 250 rpm. Cells were then centrifuged (2,200 x g for 10 min) and resuspended in 0.5 volumes of 10 mM MgSO₄. 500 μ l of this *E. coli* culture was mixed with a phage suspension containing 25,000 amplified lambda phage particles and incubated at 37°C for 15 min. To this mixture 6.5 ml of NZCYM top agarose (maintained at 60°C) (see Chart) was added and plated on 80 - 100 ml NCZYM agar (see Chart) present in a 150 mm petridish. Phage were allowed to propagate overnight at 37°C to obtain discrete plaques. After overnight growth plates were stored in a refrigerator for 1-2 hr before plaque lifts were performed.

2) Plaque Lift and DNA Hybridizations

Magna LiftTM nylon membranes (Micron Separations, Inc., Westborough, MA) were placed on the agar surface in complete contact with λ plaques and transfer of plaques to nylon membranes was allowed to proceed for 5 min at RT. After plaque transfer the membrane was placed on 2 sheets of Whatman 3MTM (Whatman, Hillsboro, OR) filter paper saturated with a 0.5 N NaOH, 1.0 M NaCl solution and left for 10 min at RT to denature DNA. Excess denaturing solution was removed by blotting briefly on dry Whatman 3M paper. Membranes were then transferred to 2 sheets of Whatman 3MTM paper saturated with 0.5 M Tris-HCl (pH 8.0), 1.5 M NaCl and left for 5 min to neutralize. Membranes were then briefly washed in 200 - 500 ml of 2 X SSC, dried by air and baked for 30 - 40 min at 80°C. The membranes were then probed with labelled DNA.

Membranes were prewashed with a 200 - 500 ml solution of 5 X SSC, 0.5% SDS, 1 mM EDTA (pH 8.0) for 1 - 2 hr at 42°C with shaking (60 rpm) to get rid of bacterial debris from the membranes. The membranes were prehybridized for 1 - 2 hr at 42°C with (in a volume equivalent to 0.125 - 0.25 ml/cm² of membrane) ECL Gold™ buffer (Amersham) containing 0.5 M NaCl and 5% blocking reagent. DNA fragments that were used as probes were purified from agarose gel using a QIAEX II™ gel extraction kit (Qiagen Inc., Chatsworth, CA) according to manufacturers protocol and labeled using an Amersham ECL™ direct nucleic acid labeling kit (Amersham). Labeled DNA (5 - 10 ng/ml hybridization solution) was added to the prehybridized membranes and the hybridization was allowed to proceed overnight. The following day membranes were washed with shaking (60 rpm) twice at 42°C for 20 min each time in (in a volume equivalent to 2 ml/cm² of membrane) a buffer containing either 0.1 (high stringency) or 0.5 (low stringency) X SSC, 0.4% SDS and 360 g/l urea. This was followed by two 5 min washes at room temperature in (in a volume equivalent to 2 ml/cm² of membrane) 2 X SSC. Hybridization signals were generated using the ECL™ nucleic acid detection reagent and detected using Hyperfilm ECL™ (Amersham).

Agar plugs which contained plaques corresponding to positive signals on the X-ray film were taken from the master plates using the broad-end of Pasteur pipet. Plaques were selected by aligning the plates with the x-ray film. At this stage, multiple plaques were generally taken. Phage particles were eluted from the agar plugs by soaking in 1 ml SM buffer (Sambrook et al., *supra*) overnight. The phage eluate was then diluted and plated with freshly grown *E. coli* XL1Blue-MRF' cells to obtain 100 - 500 plaques per 85 mm NCZYM agar plate. Plaques were transferred to Magna Lift nylon membranes as before and probed again using the same probe. Single well-isolated plaques corresponding to signals on X - ray film were picked by removing agar plugs and eluting the phage by soaking overnight in 0.5 ml SM buffer.

25

C. Conversion of λ Clones to Plasmid Form

The lambda clones isolated were converted to plasmid form for further analysis. Conversion from the plaque to the plasmid form was accomplished by infecting the plaques into *E. coli* strain BM25.8. The *E. coli* strain was grown overnight at 31°C, 250 rpm in LB broth containing 10 mM MgSO₄ and 0.2% maltose until the OD₆₀₀ reached 1.1 - 1.4. Ten milliliters of the overnight culture was removed and mixed with 100 μ l of 1 M MgCl₂. A 200 μ l volume of

30

cells was removed, mixed with 150 μ l of eluted phage suspension and incubated at 31°C for 30 min. LB broth (400 μ l) was added to the tube and incubation was continued at 31°C for 1 hr with shaking, 250 rpm. 1 - 10 μ l of the infected cell suspension was plated on LB agar containing 100 μ g/ml ampicillin (Sigma, St. Louis, MO). Well-isolated colonies were picked
5 and grown overnight in 5 ml LB broth containing 100 μ g/ml ampicillin at 37°C, 250 rpm. Plasmid DNA was isolated from these cultures and analyzed. To convert the λ ZAP Express™ vector to plasmid form *E. coli* strains XL1Blue-MRF' and XLOR were used. The conversion was performed according to the manufacturer's (Stratagene) protocols for single-plaque excision.

10

EXAMPLE 4

Transformation of *C. tropicalis* H5343 *ura*

A. Transformation of *C. tropicalis* H5343 by Electroporation

5 ml of YEPD was inoculated with *C. tropicalis* H5343 *ura*- from a frozen
15 stock and incubated overnight on a New Brunswick shaker at 30°C and 170 rpm. The next day, 10 μ l of the overnight culture was inoculated into 100 ml YEPD and growth was continued at 30°C, 170 rpm. The following day the cells were harvested at an OD_{600} of 1.0 and the cell pellet was washed one time with sterile ice-cold water. The cells were resuspended in ice-cold sterile 35 % Polyethylene glycol (4,000 MW) to a density of 5×10^8 cells/ml. A 0.1 ml volume of
20 cells were utilized for each electroporation. The following electroporation protocol was followed: 1.0 μ g of transforming DNA was added to 0.1 ml cells, along with 5 μ g denatured, sheared calf thymus DNA and the mixture was allowed to incubate on ice for 15 min. The cell solution was then transferred to an ice-cold 0.2 cm electroporation cuvette, tapped to make sure the solution was on the bottom of the cuvette and electroporated. The cells were electroporated
25 using an Invitrogen electroporator (Carlsbad, CA) at 450 Volts, 200 Ohms and 250 μ F. Following electroporation, 0.9 ml SOS media (1M Sorbitol, 30% YEPD, 10 mM $CaCl_2$) was added to the suspension. The resulting culture was grown for 1 hr at 30°C, 170 rpm. Following the incubation, the cells were pelleted by centrifugation at 1500 x g for 5 min. The electroporated cells were resuspended in 0.2 ml of 1M sorbitol and plated on synthetic complete
30 media minus uracil (SC - uracil) (Nelson, *supra*). In some cases the electroporated cells were

plated directly onto SC - uracil. Growth of transformants was monitored for 5 days. After three days, several transformants were picked and transferred to SC-uracil plates for genomic DNA preparation and screening.

5 **B. Transformation of *C. tropicalis* Using Lithium Acetate**

The following protocol was used to transform *C. tropicalis* in accordance with the procedures described in *Current Protocols in Molecular Biology*, Supplement 5, 13.7.1 (1989), incorporated herein by reference.

5 ml of YEPD was inoculated with *C. tropicalis* H5343 *ura-* from a frozen stock
10 and incubated overnight on a New Brunswick shaker at 30°C and 170 rpm. The next day, 10 µl of the overnight culture was inoculated into 50 ml YEPD and growth was continued at 30°C, 170 rpm. The following day the cells were harvested at an OD₆₀₀ of 1.0. The culture was transferred to a 50 ml polypropylene tube and centrifuged at 1000 X g for 10 min. The cell pellet was resuspended in 10 ml sterile TE (10mM Tris-Cl and 1mM EDTA, pH 8.0). The cells were again
15 centrifuged at 1000 X g for 10 min and the cell pellet was resuspended in 10 ml of a sterile lithium acetate solution [LiAc (0.1 M lithium acetate, 10 mM Tris-Cl, pH 8.0, 1 mM EDTA)]. Following centrifugation at 1000 X g for 10 min., the pellet was resuspended in 0.5 ml LiAc. This solution was incubated for one hour at 30°C while shaking gently at 50 rpm. A 0.1 ml aliquot of this suspension was incubated with 5 µg of transforming DNA at 30°C with no
20 shaking for 30 min. A 0.7 ml PEG solution (40 % wt/vol polyethylene glycol 3340, 0.1 M lithium acetate, 10 mM Tris-Cl, pH 8.0, 1 mM EDTA) was added and incubated at 30°C for 45 min. The tubes were then placed at 42°C for 5 min. A 0.2 ml aliquot was plated on synthetic complete media minus uracil (SC - uracil) (Kaiser et al. *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory Press, USA, 1994, incorporated herein by reference). Growth of
25 transformants was monitored for 5 days. After three days, several transformants were picked and transferred to SC-uracil plates for genomic DNA preparation and screening.

EXAMPLE 5

Plasmid DNA Isolation

30 Plasmid DNA were isolated from *E. coli* cultures using Qiagen plasmid isolation kit (Qiagen Inc., Chatsworth, CA) according to manufacturer's instructions.

EXAMPLE 6**DNA Sequencing and Analysis**

DNA sequencing was performed at Sequetech Corporation (Mountain View, CA) using Applied Biosystems automated sequencer (Perkin Elmer, Foster City, CA). DNA sequences were analyzed with MacVector and GeneWorks software packages (Oxford Molecular Group, Campbell, CA).

EXAMPLE 7**PCR Protocols**

PCR amplification was carried out in a Perkin Elmer Thermocycler using the AmpliTaqGold enzyme (Perkin Elmer Cetus, Foster City, CA) kit according to manufacturer's specifications. Following successful amplification, in some cases, the products were digested with the appropriate enzymes and gel purified using QiaexII (Qiagen, Chatsworth, CA) as per manufacturer instructions. In specific cases the Ultima Taq polymerase (Perkin Elmer Cetus, Foster City, CA) or the Expand Hi-Fi Taq polymerase (Boehringer Mannheim, Indianapolis, IN) were used per manufacturer's recommendations or as defined in Table 3.

Table 3. PCR amplification conditions used with different primer combinations.

PRIMER COMBINATION	Taq	TEMPLATE DENATURING CONDITION	ANNEALING TEMP/TIME	EXTENSION TEMP/TIME	CYCLE Number
3674-41-1/ 41-2/ 41-4 + 3674-41-4	Ampli-Taq Gold	94 C/30 sec	55 C/30 sec	72 C/1 min	30
URA Primer 1a URA Primer 1b	Ampli-Taq Gold	95 C/1 min	70 C/1 min	72 C/2 min	35
URA Primer 2a URA Primer 2b	Ampli-Taq Gold	95 C/1 min	70 C/1 min	72 C/2 min	35
CYP2A#1 CYP2A#2	Ampli-Taq Gold	95 C/1 min	70 C/1 min	72 C/2 min	35
CYP3A#1 CYP3A#2	Ultima Taq	95 C/1 min	70 C/1 min	72 C/1 min	30
CPR B#1	Expand	94 C/15 sec	50 C/30 sec	68 C/3 min	10
CPR B#2	Hi-Fi Taq	94 C/15 sec	50 C/30 sec	68 C/3 min +20 sec/cycle	15

CYP5A#1	Expand	94 C/15 s c	50 C/30 sec	68 C/3 min	10
CYP5A#2	Hi-Fi	94 C/15 sec	50 C/30 sec	68 C/3 min	15
	Taq			+20 sec/cycle	

5 Table 4 below contains a list of primers (SEQ ID NOS: 1-35) used for PCR amplification to construct gene integration vectors or to generate probes for gene detection and isolation.

10 Table 4. Primer table for PCR amplification to construct gene integration vectors, to generate probes for gene isolation and detection and to obtain DNA sequence of constructs. (A- deoxyadenosine triphosphate [dATP], G- deoxyguanosine triphosphate [dGTP], C- deoxycytosine triphosphate [dCTP], T- deoxythymidine triphosphate [dTTP], Y- dCTP or dTTP, R- dATP or dGTP, W- dATP or dTTP, M- dATP or dCTP, N- dATP or dCTP or dGTP or dTTP).

15	Target gene(s)	Patent Primer Name	Lab Primer Name	Sequence (5' to 3')	PCR Product Size
	CYP52A2A	CYP2A#1	3659-72M	CCTTAATTAAATGCACGAAGCGGAGA TAAAAG (SEQ ID NO: 1)	2230 bp
		CYP2A#2	3659-72N	CCITTAATTAAGCATAAGCTTGCTCGAG TCT (SEQ ID NO: 2)	
20	CYP52A3A	CYP3A#1	3659-72O	CCTTAATTAAACGCAATGGGAACATG GAGTG (SEQ ID NO: 3)	2154 bp
		CYP3A#2	3659-72P	CCITTAATTAATCGCACTACGGTTATTG GTATCAG (SEQ ID NO: 4)	
	CYP52A5A	CYP5A#1	3659-72K	CCTTAATTAATCAAAGTACGTTACAGG GG (SEQ ID NO: 5)	3298 bp
		CYP5A#2	3659-72L	CCITTAATTAAGGCAGACAACAACCTTG GCAAAGTC (SEQ ID NO: 6)	
25	CPRB	CPRB#1	3698-20A	CCTTAATTAAGAGGTCGTTGTTGAGT TTTC (SEQ ID NO: 7)	3266 bp
		CPRB#2	3698-20B	CCITTAATTAATTGATAATGACGTTGCG GG (SEQ ID NO: 8)	
	URA3A	URA Primer 1a	3698-7C	AGGCGCGCCGAGTCCAAAAAGACC AACCTCTG (SEQ ID NO: 9)	956 bp
30		URA Primer 1b	3698-7D	CCTTAATTAATACGTGGATACCTTCAA GCAAAGTG (SEQ ID NO: 10)	

5	URA3A	URA Primer 2a	3698-7A	CCTTAATTAAGCTCAGAGTTTGGGA TTTTCGAG (SEQ ID NO: 11)	750 bp
		URA Primer 2b	3698-7B	GGGTTTAAACCGCAGAGTTGGTCTT TTTGGACTC (SEQ ID NO: 12)	
				GGGTTTAAAC - Pme I restriction site (SEQ ID NO: 13)	
				AGGCGCGCC - AseI restriction site (SEQ ID NO: 14)	
				CCTTAATTA - PacI restriction site (SEQ ID NO: 15)	
10	CPR	FMN1	3674-41-1	TCYCAAACWGGTACWGCWGAA (SEQ ID NO: 16)	
	CPR	FMN2	3674-41-2	GGTTGGGTAAAYTCWACTTAT (SEQ ID NO: 17)	
	CPR	FAD	3674-41-3	CGTTATTAYTCYATTCTTC (SEQ ID NO: 18)	
	CPR	NADPH	3674-41-4	GCMACACCRGTACCTGGACC (SEQ ID NO: 19)	
	CPR	PRK1.F3	PRK1.F3	ATCCCAATCGTAATCAGC (SEQ ID NO: 20)	
15	CPR	PRK1.F5	PRK1.F5	ACTTGTCTTCGTTTAGCA (SEQ ID NO: 21)	
	CPR	PRK4.R20	PRK4.R20	CTACGTCTGTGGTGATGC (SEQ ID NO: 22)	
	CYP	UCup1	UCup1	CGNGAYACNACNGCNGG (SEQ ID NO: 23)	
	CYP	UCup2	UCup2	AGRGAYACNACNGCNGG (SEQ ID NO: 24)	
	CYP	UCdown1	UCdown1	AGNGCRAAYTYGTGNCC (SEQ ID NO: 25)	
20	CYP	UCdown2	UCdown2	YAANGCRAAYTYGTGNCC (SEQ ID NO: 26)	
	CYP	HemeB1	HemeB1	ATTCAACGGTGGTCCAAGAATCTGTT TGG (SEQ ID NO: 27)	
	CYP	2,3,5P	2,3,5P	GAGCTATGTTGAGACCACAGTTTGC (SEQ ID NO: 28)	
	CYP	2,3,5M	2,3,5M	CTTCAGTTAAAGCAAATTGTTTGCC (SEQ ID NO: 29)	
	pTriplEx vector	Triplex5'	Triplex5'	CTCGGGAAGCGGCCATTGTGTTGG (SEQ ID NO: 30)	
25	pTriplEx vector	Triplex3'	Triplex3'	TAATACGACTCACTATAGGGCGAAT TGGC (SEQ ID NO: 31)	
	CYP	Cyp52a	Cyp52a	TGRYTCAAACCATCTYTCTGG (SEQ ID NO: 32)	
	CYP	Cyp52b	Cyp52b	GGACCGGCGTTAAAGGG (SEQ ID NO: 33)	
	CYP	Cyp52c	Cyp52c	CATAGTCGWATYATGCTTAGACC (SEQ ID NO: 34)	
	CYP	Cyp52d	Cyp52d	GGACCACCATGAATGG (SEQ ID NO: 35)	

30

EXAMPLE 8**Yeast Colony PCR Procedure for Confirmation of Gene
Integration into the Genome of *C. tropicalis***

5 Single yeast colonies were removed from the surface of transformation plates, suspended in 50 μ l of spheroplasting buffer (50mM KCl, 10mM Tris-HCl, pH 8.3, 1.0 mg/ml Zymolyase, 5% glycerol) and incubated at 37°C for 30 min. Following incubation, the solution was heated for 10 min at 95°C to lyse the cells. Five μ l of this solution was used as a template in PCR. Expand Hi-Fi *Taq* polymerase (Boehringer Mannheim, Indianapolis, IN) was used in PCR
10 coupled with a gene-specific primer (gene to be integrated) and a *URA3* primer. If integration did occur, amplification would yield a PCR product of predicted size confirming the presence of an integrated gene.

EXAMPLE 9**15 Fermentation Method for Gene Induction Studies**

 A fermentor was charged with a semi-synthetic growth medium having the composition 75 g/l glucose (anhydrous), 6.7 g/l Yeast Nitrogen Base (Difco Laboratories), 3 g/l yeast extract, 3 g/l ammonium sulfate, 2 g/l monopotassium phosphate, 0.5 g/l sodium chloride. Components were made as concentrated solutions for autoclaving then added to the fermentor
20 upon cooling: final pH approximately 5.2. This charge was inoculated with 5-10% of an overnight culture of *C. tropicalis* ATCC 20962 prepared in YM medium (Difco Laboratories) as described in the methods of Examples 17 and 20 of US Patent 5,254,466, which is incorporated herein by reference. *C. tropicalis* ATCC 20962 is a POX 4 and POX 5 disrupted *C. tropicalis* ATCC 20336. Air and agitation were supplied to maintain the dissolved oxygen at greater than
25 about 40% of saturation versus air. The pH was maintained at about 5.0 to 8.5 by the addition of 5N caustic soda on pH control. Both a fatty acid feedstream (commercial oleic acid in this example) having a typical composition: 2.4% C₁₄; 0.7% C_{14:1}; 4.6% C₁₆; 5.7% C_{16:1}; 5.7% C_{17:1}; 1.0% C₁₈; 69.9% C_{18:1}; 8.8% C_{18:2}; 0.30% C_{18:3}; 0.90% C_{20:1} and a glucose co-substrate feed were added in a feedbatch mode beginning near the end of exponential growth. Caustic was added on
30 pH control during the bioconversion of fatty acids to diacids to maintain the pH in the desired range. Typically, samples for gene induction studies were collected just prior to starting the fatty acid feed and over the first 10 hours of bioconversion. Determination of fatty acid and diacid

content was determined by a standard methyl ester protocol using gas liquid chromatography (GLC). Gene induction was measured using the QC-RT-PCR protocol described in this application.

5

EXAMPLE 10

RNA Preparation

The first step of this protocol involves the isolation of total cellular RNA from cultures of *C. tropicalis*. The cellular RNA was isolated using the Qiagen RNeasy Mini Kit (Qiagen Inc., Chatsworth, CA) as follows: 2 ml samples of *C. tropicalis* cultures were collected from the fermentor in a standard 2 ml screw capped Eppendorf style tubes at various times before and after the addition of the fatty acid or alkane substrate. Cell samples were immediately frozen in liquid nitrogen or a dry-ice/alcohol bath after their harvesting from the fermentor. To isolate total RNA from the samples, the tubes were allowed to thaw on ice and the cells pelleted by centrifugation in a microfuge for 5 minutes (min) at 4°C and the supernatant was discarded while keeping the pellet ice-cold. The microfuge tubes were filled 2/3 full with ice-cold Zirconia/Silica beads (0.5 mm diameter, Biospec Products, Bartlesville, OK) and the tube filled to the top with ice-cold RLT* lysis buffer (* buffer included with the Qiagen RNeasy Mini Kit). Cell rupture was achieved by placing the samples in a mini bead beater (Biospec Products, Bartlesville, OK) and immediately homogenized at full speed for 2.5 min. The samples were allowed to cool in a ice water bath for 1 minute and the homogenization/cool process repeated two more times for a total of 7.5 min homogenization time in the beadbeater. The homogenized cells samples were microfuged at full speed for 10 min and 700 µl of the RNA containing supernatant removed and transferred to a new eppendorf tube. 700 µl of 70% ethanol was added to each sample followed by mixing by inversion. This and all subsequent steps were performed at room temperature. Seven hundred microliters of each ethanol treated sample were transferred to a Qiagen RNeasy spin column, followed by centrifugation at 8,000 x g for 15 sec. The flow through was discarded and the column reloaded with the remaining sample (700 µl) and re-centrifuged at 8,000 x g for 15 sec. The column was washed once with 700 µl of buffer RW1*, and centrifuged at 8,000 x g for 15 sec and the flow through discarded. The column was placed in a new 2 ml collection tube and washed with 500 µl of RPE* buffer and the flow through discarded. The RPE* wash was repeated with centrifugation at 8,000 x g for 2 min and the flow through

10

15

20

25

30

discarded. The spin column was transferred to a new 1.5 ml collection tube and 100 μ l of RNase free water added to the column followed by centrifugation at 8,000 x g for 15 seconds. An additional 75 μ l of RNase free water was added to the column followed by centrifugation at 8,000 x g for 2 min. RNA eluted in the water flow through was collected for further purification.

- 5 The RNA eluate was then treated to remove contaminating DNA. Twenty microliters of 10X DNase I buffer (0.5 M tris (pH 7.5), 50 mM CaCl_2 , 100 mM MgCl_2), 10 μ l of RNase-free DNase I (2 Units/ μ l, Ambion Inc., Austin, Texas) and 40 units Rnasin (Promega Corporation, Madison, Wisconsin) were added to the RNA sample. The mixture was then incubated at 37°C for 15 to 30 min. Samples were placed on ice and 250 μ l Lysis buffer RLT* and 250 μ l ethanol (200 proof) added. The samples were then mixed by inversion. The samples were transferred to Qiagen RNeasy spin columns and centrifuged at 8,000 x g for 15 sec and the flow through discarded. Columns were placed in new 2 ml collection tubes and washed twice with 500 μ l of RPE* wash buffer and the flow through discarded. Columns were transferred to new 1.5 ml eppendorf tubes and RNA was eluated by the addition of 100 μ l of DEPC treated water followed by centrifugation at 8,000 x g for 15 sec. Residual RNA was collected by adding an additional 50 μ l of RNase free water to the spin column followed by centrifugation at full speed for 2 min. 10 μ l of the RNA preparation was removed and quantified by the ($A_{260/280}$) method. RNA was stored at -70°C. Yields were found to be 30-100 μ g total RNA per 2.0 ml of fermentation broth.

20

EXAMPLE 11

Quantitative Competitive Reverse Transcription Polymerase Chain Reaction (QC-RT-PCR) Protocol

- 25 QC-RT-PCR is a technique used to quantitate the amount of a specific RNA in a RNA sample. This technique employs the synthesis of a specific DNA molecule that is complementary to an RNA molecule in the original sample by reverse transcription and its subsequent amplification by polymerase chain reaction. By the addition of various amounts of a competitor RNA molecule to the sample one can determine the concentration of the RNA molecule of interest (in this case the mRNA transcripts of the *CYP* and *CPR* genes). The levels of specific mRNA transcripts were assayed over time in response to the addition of fatty acid
- 30

and/or alkane substrates to the growth medium of fermentation grown *C. tropicalis* cultures for the identification and characterization of the genes involved in the oxidation of these substrates. This approach can be used to identify the *CYP* and *CPR* genes involved in the oxidation of any given substrate based upon their transcriptional regulation.

5

A. Primer Design

The first requirement for QC-RT-PCR is the design of the primer pairs to be used in the reverse transcription and subsequent PCR reactions. These primers need to be unique and specific to the gene of interest. As there is a family of genetically similar *CYP* genes present in *C. tropicalis* 20336, care had to be taken to design primer pairs that would be discriminating and only amplify the gene of interest, in this example the *CYP52A5* gene. In this manner, unique primers directed to substantially non-homologous (aka variable) regions within target members of a gene family are constructed. What constitutes substantially non-homologous regions is determined on a case by case basis. Such unique primers should be specific enough to anneal the non-homologous region of the target gene without annealing to other non-target members of the gene family. By comparing the known sequences of the members of a gene family, non-homologous regions are identified and unique primers are constructed which will anneal to those regions. It is contemplated that non-homologous regions herein would typically exhibit less than about 85% homology but can be more homologous depending on the positions which are conserved and stringency of the reaction. After conducting PCR, it may be helpful to check the reaction product to assure it represents the unique target gene product. If not, the reaction conditions can be altered in terms of stringency to focus the reaction to the desired target. Alternatively a new primer or new non-homologous region can be chosen. Due to the high level of homology between the genes of the *CYP52A* family, the most variable 5 prime region of the *CYP52A5* coding sequence was targeted for the design of the primer pairs. In Figure 3, a portion of the 5 prime coding region for the *CYP52A5A* (SEQ ID NO: 36) allele of *C. tropicalis* 20336 is shown. The boxed sequences in Figure 3 are the sequences of the forward and backwards primers (SEQ ID NOS: 47 and 48) used to quantitate expression of both alleles of this gene. The actual reverse primer (SEQ ID NO: 48) contains one less adenine than that shown in Figure 3. Primers used to measure the expression of specific *C. tropicalis* 20336 genes using the QC-RT-PCR protocol are listed in Table 5 (SEQ ID NOS: 37-58).

Table 5. Primer used to measure *C. tropicalis* gene expression in the QC-RT-PCR reactions.

	Primer Name	Direction	Target	Sequence
5	3737-89F	F	<i>CYP52A1A</i>	CCGATGAAGTTTTCGACGAGTACCC (SEQ ID NO: 37)
	3737-89B	B	<i>CYP52A1A</i>	AAGGCTTTAACGTGTCCAATCTGGTC (SEQ ID NO: 38)
	alk2aF1	F	<i>CYP52A2A</i>	ATTATCGCCACATACTTCACCAAATGG (SEQ ID NO: 39)
	alk2aB5	B	<i>CYP52A2A</i>	CGAGATCGTGGATACGCTGGAGTG (SEQ ID NO: 40)
10	7581-178-3	F	<i>CYP52A3A</i>	GCCACTCGGTAACTTTGTCAAGGAC (SEQ ID NO: 41)
	7581-178-4	B	<i>CYP52A3A</i>	CATTGAACTGAGTAGCCAAAACAGCC (SEQ ID NO: 42)
	3737-50F	F	<i>CYP52A3A</i> & <i>CYP52A3B</i>	CCTACGTTTGGTATCGCTACTCCGTTG (SEQ ID NO: 43)
	3737-50B	B	<i>CYP52A3A</i> & <i>CYP52A3B</i>	TTCCAGCCAGCACCGTCCAAG (SEQ ID NO: 44)
15	3737-175F	F	<i>CYP52D4A</i>	GCAGAGCCGATCTATGTTGCGTCC (SEQ ID NO: 45)
	3737-175B	B	<i>CYP52D4A</i>	TCATTGAATGCTTCCAGGAACCTCG (SEQ ID NO: 46)
	7581-97-F	F	<i>CYP52A5A</i> & <i>CYP52A5B</i>	AAGAGGGCAGGGCTCAAGAG (SEQ ID NO: 47)
	7581-97-M	B	<i>CYP52A5A</i> & <i>CYP52A5B</i>	TCCATGTGAAGATCCCATCAC (SEQ ID NO: 48)
20	4P-2	F	<i>CYP52A8A</i>	CTTGAAGCCGTGTTGAACG (SEQ ID NO: 49)
	4M-1	B	<i>CYP52A8A</i>	CAGGATTGTCTGAGTTGCCG (SEQ ID NO: 50)
	3737-52F	F	<i>POX4A</i> & <i>POX4B</i>	CCATTGCCTTGAGATACGCCATTGGTAG (SEQ ID NO: 51)
	3737-52B	B	<i>POX4A</i> & <i>POX4B</i>	AGCCTTGGTGTCTGTTCTTTCAACGG (SEQ ID NO: 52)
25	3737-53F	F	<i>POX5A</i>	TGGGTTTGTGTTTCTCCTGTGTCCG (SEQ ID NO: 53)
	3737-53B	B	<i>POX5A</i>	CCTTGACCTTCAATCTGGCGTAGACG (SEQ ID NO: 54)
	F33	F	<i>CPRA</i>	GGTTTGCTGAATACGCTGAAGGTGATG (SEQ ID NO: 55)
	B63	B	<i>CPRA</i>	TGGAGCTGAACAACCTCTCTCGTCTCGG (SEQ ID NO: 56)
25	3737-133F	F	<i>CPRA</i> & <i>CPRB</i>	TTCTCAACACGGACAGCGG (SEQ ID NO: 57)
	3737-133B	B	<i>CPRA</i> & <i>CPRB</i>	AGTCAACCAGGTGTGGAACCTCGTC (SEQ ID NO: 58)

F=Forward B=Backward

B. Design and Synthesis of the Competitor DNA Template

The competitor RNA is synthesized *in vitro* from a competitor DNA template that has the T7 polymerase promoter and preferably carries a small deletion of e.g., about 10 to 25 nucleotides relative to the native target RNA sequence. The DNA template for the *in-vitro* synthesis of the competitor RNA is synthesized using PCR primers that are between 46 and 60 nucleotides in length. In this example, the primer pairs for the synthesis of the *CYP52A5* competitor DNA are shown in Tables 6 and 7 (SEQ ID NOS: 59 AND 60).

10 **Table 6.** Forward and Reverse primers used to synthesize the competitor RNA template for the QC-RT-PCR measurement of *CYP52A5A* gene expression.

Forward Primer	<i>CYP52A5A</i>	GGATCCTAATACGACTCACTATAGGGAGGA AGAGGGCAGGGCTCAAGAG (SEQ ID NO: 59)
Reverse Primer	<i>CYP52A5A</i>	TCCATGTGAAGATCCCATCACGAGTGTGCC TCTTGCCCAAAG (SEQ ID NO: 60)

15 **Table 7.** Primers for the synthesis of the QC-RT-PCR competitor RNA templates

Primer Name	Direction	Target	Sequence 5'-3'
3737-89C	F	<i>CYP52A1A</i>	GGATCCTAATACGACTCACTATAGGGAGGCCGATG AAGTTTTGACGAGTACCC (SEQ ID NO: 61)
3737-89D	B	<i>CYP52A1A</i>	AAGGCTTTAACGTGTCCAATCTGGTC AACATAGCTCTGGAGTGCTTCCAACC (SEQ ID NO: 62)
7581-137-A	F	<i>CYP52A2A</i>	GGATCCTAATACGACTCACTATAGGGAGGATTATC GCCACATACTTCACCAAATGG (SEQ ID NO: 63)
7581-137-B	B	<i>CYP52A2A</i>	CGAGATCGTGGATACGCTGGAGTGCGTCGCTCTTC TTCTTCAACAATTCAAG (SEQ ID NO: 64)
7581-137-D	B	<i>CYP52A3A</i>	CATTGAACTGAGTAGCCAAAACAGCCCATGGTTTC AATCAATGGGAGGC (SEQ ID NO: 65)
7581-137-C	F	<i>CYP52A3A</i>	GGATCCTAATACGACTCACTATAGGGAGGGCCACT CGGTAACCTTGTGAGGGAC (SEQ ID NO: 66)

3737-50-D	F	CYP52A3A & CYP52A3B	GGATCCTAATACGACTCACTATAGGGAGGCCTACG TTTGGTATCGCTACTCCGTTG (SEQ ID NO: 67)
3737-50-C	B	CYP52A3A & CYP52A3B	TTTCCAGCCAGCACCGTCCAAGCAACAAGGAGTAC AAGAAATCGTGTC (SEQ ID NO: 68)
3737-175C	F	CYP52D4A	GGATCCTAATACGACTCACTATAGGGAGGGCAGAG CCGATCTATGTTGCGTCC (SEQ ID NO: 69)
3737-175D	B	CYP52D4A	TCATTGAATGCTTCCAGGAACCTCGCCACATCCATC GAGAACCGG (SEQ ID NO: 70)
5 7581-97-A	F	CYP52A5A & CYP52A5B	GGATCCTAATACGACTCACTATAGGGAGGAAGAGG GCAGGGCTCAAGAG (SEQ ID NO: 59)
7581-97-B	B	CYP52A5A & CYP52A5B	TCCATGTGAAGATCCCATCACGAGTGTGCCTCTTGC CCAAAG (SEQ ID NO: 60)
4P-2/T7	F	CYP52A8A	GGATCCTAATACGACTCACTATAGGGAGGCTTGAA GGCCGTGTTGAACG (SEQ ID NO: 71)
4M-3/4M-1	B	CYP52A8A	CAGGATTGTCTGAGTTGCCGCCTGATCAAGATAG GATCCTTGCCG (SEQ ID NO: 72)
3737-26-D	F	CPRA	GGATCCTAATACGACTCACTATAGGGAGGGGTTTG CTGAATACGCTGAAGGTGATG (SEQ ID NO: 73)
10 3737-26-C	B	CPRA	TGGAGCTGAACAACTCTCTCGTCTCGGGTGGTCTGA ATGGACCCTTGGTCAAG (SEQ ID NO: 74)
3737-133C	F	CPRA & CPRB	GGATCCTAATACGACTCACTATAGGGAGGTTCTC AACACGGACAGCGG (SEQ ID NO: 75)
3737-133D	B	CPRA & CPRB	AGTCAACCAGGTGTGGAACCTCGTCGGTGGCAACAA TGAAAAACACCAAG (SEQ ID NO: 76)
3737-52-C	F	POX4A & POX4B	GGATCCTAATACGACTCACTATAGGGAGGCCATTG CCTTGAGATACGCCATTGGTAG (SEQ ID NO: 77)
3737-52-D	B	POX4A & POX4B	AGCCTTGGTGTCTGTTCTTTCAACGGAAGGTGGTCT CGATGGTGTGTTCAACC (SEQ ID NO: 78)
15 3737-53-C	F	POX5A	GGATCCTAATACGACTCACTATAGGGAGGTTGGGT TTGTTTGTTCCTGTGTCCG (SEQ ID NO: 79)
3737-53-D	B	POX5A	CCTTTGACCTTCAATCTGGCGTAGACGCAGCACCA CCGATCCACCACTTG (SEQ ID NO: 80)

F=Forward B=Backword

The forward primer (SEQ ID NO: 59) contains the T7 promoter consensus sequence "GGATCCTAATACGA CTCACTATAGGG AGG" fused to the primer 7581-97-F sequence (SEQ ID NO: 47). The Reverse Primer (SEQ ID NO: 60) contains the sequence of primer 7581-97M (SEQ ID NO: 48) followed by the 20 bases of upstream sequence with a 18 base pair deletion between the two blocks of the *CYP52A5* sequence. The forward primer was used with the corresponding reverse primer to synthesize the competitor DNA template. The primer pairs were combined in a standard *Taq* Gold polymerase PCR reaction according to the manufacturer's recommended conditions (Perkin-Elmer/Applied Biosystems, Foster City, CA). The PCR reaction mix contained a final concentration of 250 nM each primer and 10 ng *C. tropicalis* chromosomal DNA for template. The reaction mixture was placed in a thermocycler for 25 to 35 cycles using the highest annealing temperature possible during the PCR reactions to assure a homogeneous PCR product (in this case 62°C). The PCR products were either gel purified or filtered purified to remove un-incorporated nucleotides and primers. The competitor template DNA was then quantified using the ($A_{260/280}$) method. Primers used in QC-RT-PCR experiments for the synthesis of various competitive DNA templates are listed in Table 7 (SEQ ID NOS: 61-80).

C. Synthesis of the Competitor RNA

Competitor template DNA was transcribed *In-Vitro* to make the competitor RNA using the Megascript T7 kit from Ambion Biosciences (Ambion Inc., Austin, Texas). 250 nanograms (ng) of competitor DNA template and the *in-vitro* transcription reagents are mixed according to the directions provided by the manufacturer. The reaction mixture was incubated for 4 hours at 37°C. The resulting RNA preparations were then checked by gel electrophoresis for the conditions giving the highest yields and quality of competitor RNA. This often required optimization according to the manufacturer's specifications. The DNA template was then removed using DNase I as described in the Ambion kit. The RNA competitor was then quantified by the ($A_{260/280}$) method. Serial dilution's of the RNA (1 ng/ μ l to 1 femtogram (fg)/ μ l) were made for use in the QC-RT-PCR reactions and the original stocks stored at -70°C.

30

D. QC-RT-PCR Reactions

QC-RT-PCR reactions were performed using rTth polymerase from Perkin-Elmer(Perkin-Elmer/Applied Biosystems, Foster City, CA) according to the manufacturer's recommended conditions. The reverse transcription reaction was performed in a 10 µl volume with a final concentrations of 200 µM for each dNTP, 1.25 units rTth polymerase, 1.0 mM MnCl₂, 1X of the 10X buffer supplied with the Enzyme from the manufacturer, 100 ng of total RNA isolated from a fermentor grown culture of *C. tropicalis* and 1.25 µM of the appropriate reverse primer. To quantitate *CYP52A5* expression in *C. tropicalis* an appropriate reverse primer was 7581-97M (SEQ ID NO: 48). Several reaction mixes were prepared for each RNA sample characterized. To quantitate *CYP52A5* expression a series of 8 to 12 of the previously described QC-RT-PCR reaction mixes were aliquoted to different reaction tubes. To each tube 1 µl of a serial dilution containing from 100 pg to 100 fg *CYP52A5* competitor RNA per µl was added bringing the final reaction mixtures up to the final volume of 10 µl. The QC-RT-PCR reaction mixtures were mixed and incubated at 70°C for 15 min according to the manufacturer's recommended times for reverse transcription to occur. At the completion of the 15 minute incubation, the sample temperature was reduced to 4°C to stop the reaction and 40 µl of the PCR reaction mix added to the reaction to bring the total volume up to 50 µl. The PCR reaction mix consists of an aqueous solution containing 0.3125 µM of the forward primer 7581-97F (SEQ ID NO: 47), 3.125 mM MgCl₂, and 1X chelating buffer supplied with the enzyme from Perkin-Elmer. The reaction mixtures were placed in a thermocycler (Perkin-Elmer GeneAmp PCR System 2400, Perkin-Elmer/Applied Biosystems, Foster City, CA) and the following PCR cycle performed: 94°C for 1 min. followed by 94°C for 10 seconds followed by 58°C for 40 seconds for 17 to 22 cycles. The PCR reaction was completed with a final incubation at 58°C for 2 min followed by 4°C. In some reactions where no detectable PCR products were produced the samples were returned the thermocycler for additional cycles, this process was repeated until enough PCR products were produced to quantify using HPLC. The number of cycles necessary to produce enough PCR product is a function of the amount of the target mRNA in the 100 ng of total cellular RNA. In cultures where the *CYP52A5* gene is highly expressed there is sufficient *CYP52A5* mRNA message present and less PCR cycles (≤ 17) are required to produce quantifiable amount of PCR product. The lower the concentrations of the target mRNA present the more PCR cycles are required to produce a detectable amount of product. These QC-RT-

PCR procedures were applied to all the target genes listed in Table 5 using the respective primers indicated therein.

E. HPLC Quantification

5 Upon completion of the QC-RT-PCR reactions the samples were analyzed and quantitated by HPLC. Five to fifteen microliters of the QC-RT-PCR reaction mix was injected into a Waters Bio-Compatible 625 HPLC with an attached Waters 484 tunable detector. The detector was set to measure a wave length of 254 nm. The HPLC contained a Sarasep brand DNASep™ column (Sarasep, Inc., San Jose, CA) which was placed within the oven and the
10 temperature set for 52 °C. The column was installed according to the manufacturer's recommendation of having 30 cm. of heated PEEK tubing installed between the injector and the column. The system was configured with a Sarasep brand Guard column positioned before the injector. In addition, there was a 0.22 µm filter disk just before the column, within the oven. Two Buffers were used to create an elution gradient to resolve and quantitate the PCR products
15 from the QC-RT-PCR reactions. Buffer-A consists of 0.1 M tri-ethyl ammonium acetate (TEAA) and 5% acetonitrile (volume to volume). Buffer-B consists of 0.1 M TEAA and 25% acetonitrile (volume to volume). The QC-RT-PCR samples were injected into the HPLC and the linear gradient of 75% buffer-A/ 25% buffer-B to 45% buffer-A/ 55% B was run over 6 min at a flow rate of 0.85 ml per minute. The QC-RT-PCR product of the competitor RNA being 18
20 base pairs smaller is eluted from the HPLC column before the QC-RT-PCR product from the *CYP52A5* mRNA(U). The amount of the QC-RT-PCR products are plotted and quantitated with an attached Waters Corporation 745 data module. The log ratios of the amount of *CYP52A5* mRNA QC-RT-PCR product (U) to competitor QC-RT-PCR product (C), as measured by peak areas, was plotted and the amount of competitor RNA required to equal the amount of *CYP52A5*
25 mRNA product determined. In the case of each of the target genes listed in Table 5, the competitor RNA contained fewer base pairs as compared to the native target mRNA and eluted before the native mRNA in a manner similar to that demonstrated by *CYP52A5*. HPLC quantification of the genes was conducted as above.

30

EXAMPLE 12**Evaluation of New Strains in Shake Flasks**

The *CYP* and *CPR* amplified strains such as strains HDC10, HDC15, HDC20 and HDC23 (Table 1) and H5343 were evaluated for diacid production in shake flasks. A single colony for each strain was transferred from a YPD agar plate into 5 ml of YPD broth and grown overnight at 30°C, 250 rpm. An inoculum was then transferred into 50 ml of DCA2 medium (Chart) and grown for 24 h at 30°C, 300 rpm. The cells were centrifuged at 5000 rpm for 5 min and resuspended in 50 ml of DCA3 medium (Chart) and grown for 24 h at 30°C, 300 rpm. 3% oleic acid w/v was added after 24 h growth in DCA3 medium and the cultures were allowed to bioconvert oleic acid for 48 h. Samples were harvested and the diacid and monoacid concentrations were analyzed as per the scheme given in Figure 35. Each strain was tested in duplicate and the results shown in Table 8 represent the average value from two flasks.

Table 8. Bioconversion of oleic acid by different recombinant strains of *Candida tropicalis*

Strain	Conversion to Oleic diacid (%)	Specific Conversion (g diacid/g biomass)
H5343	41.9	0.53
HDC 10-2	50.5	0.85
HDC 15	54.4	0.85
HDC 20-1	45.1	0.72
HDC 20-2	45.3	0.58
HDC 23-2	55.2	0.84
HDC 23-3	58.8	0.89

EXAMPLE 13**Cloning and Characterization of *C. tropicalis* 20336 Cytochrome P450 Monooxygenase (*CYP*) and Cytochrome P450 NADPH Oxidoreductase (*CPR*) Genes**

To clone *CYP* and *CPR* genes several different strategies were employed. Available *CYP* amino acid sequences were aligned and regions of similarity were observed (Figure 4). These regions corresponded to described conserved regions seen in other cytochrome P450 families (Goeptar et al., *supra* and Kalb et al. *supra*). Proteins from eight eukaryotic

cytochrome P450 families share a segmented region of sequence similarity. One region corresponded to the HR2 domain containing the invariant cysteine residue near the carboxyl terminus which is required for heme binding while the other region corresponded to the central region of the I helix thought to be involved in substrate recognition (Figure 4). Degenerate
5 oligonucleotide primers corresponding to these highly conserved regions of the *CYP52* gene family present in *Candida maltosa* and *Candida tropicalis* ATCC 750 were designed and used to amplify DNA fragments of *CYP* genes from *C. tropicalis* 20336 genomic DNA. These discrete PCR fragments were then used as probes to isolate full-length *CYP* genes from the *C. tropicalis* 20336 genomic libraries. In a few instances oligonucleotide primers corresponding to
10 highly conserved regions were directly used as probes to isolate full-length *CYP* genes from genomic libraries. In the case of *CPR* a heterologous probe based upon the known DNA sequence for the *CPR* gene from *C. tropicalis* 750 was used to isolate the *C. tropicalis* 20336 *CPR* gene.

15 A. Cloning of the *CPR* Gene from *C. tropicalis* 20336

1) Cloning of the *CPRA* Allele

Approximately 25,000 phage particles from the first genomic library of *C. tropicalis* 20336 were screened with a 1.9 kb *Bam*HI-*Nde*I fragment from plasmid pCU3RED (See Picattagio et al., Bio/Technology 10:894-898 (1992), incorporated herein by reference)
20 containing most of the *C. tropicalis* 750 *CPR* gene. Five clones that hybridized to the probe were isolated and the plasmid DNA from these lambda clones was rescued and characterized by restriction enzyme analysis. The restriction enzyme analysis suggested that all five clones were identical but it was not clear that a complete *CPR* gene was present.

PCR analysis was used to determine if a complete *CPR* gene was present in any of
25 the five clones. Degenerate primers were prepared for highly conserved regions of known *CPR* genes (See Sutter et al., *J. Biol. Chem.* 265:16428-16436 (1990), incorporated herein by reference) (Figure 4). Two Primers were synthesized for the FMN binding region (FMN1, SEQ ID NO: 16 and FMN2, SEQ ID NO: 17). One primer was synthesized for the FAD binding region (FAD, SEQ ID NO: 18), and one primer for the NADPH binding region (NADPH, SEQ
30 ID NO: 19) (Table 4). These four primers were used in PCR amplification experiments using as a template plasmid DNA isolated from four of the five clones described above. The FMN (SEQ

ID NOS: 16 and 17) and FAD (SEQ ID NO: 18) primers served as forward primers and the NADPH primer (SEQ ID NO: 19) as the reverse primer in the PCR reactions. When different combinations of forward and reverse primers were used, no PCR products were obtained from any of the plasmids. However, all primer combinations amplified expected size products with a plasmid containing the *C. tropicalis* 750 *CPR* gene (positive control). The most likely reason for the failure of the primer pairs to amplify a product, was that all four of clones contained a truncated *CPR* gene. One of the four clones (pHKM1) was sequenced using the Triplex 5' (SEQ ID NO: 30) and the Triplex 3' (SEQ ID NO: 31) primers (Table 4) which flank the insert and the multiple cloning site on the cloning vector, and with the degenerate primer based upon the NADPH binding site described above. The NADPH primer (SEQ ID NO: 19) failed to yield any sequence data and this is consistent with the PCR analysis. Sequences obtained with Triplex primers were compared with *C. tropicalis* 750 *CPR* sequence using the MacVector™ program (Oxford Molecular Group, Campbell, CA). Sequence obtained with the Triplex 3' primer (SEQ ID NO: 31) showed similarity to an internal sequence of the *C. tropicalis* 750 *CPR* gene confirming that pHKM1 contained a truncated version of a 20336 *CPR* gene. pHKM1 had a 3.8 kb insert which included a 1.2 kb coding region of the *CPR* gene accompanied by 2.5 kb of upstream DNA (Figure 5). Approximately 0.85 kb of the 20336 *CPR* gene encoding the C-terminal portion of the *CPR* protein is missing from this clone.

Since the first Clontech library yielded only a truncated *CPR* gene, the second library prepared by Clontech was screened to isolate a full-length *CPR* gene. Three putative *CPR* clones were obtained. The three clones, having inserts in the range of 5-7 kb, were designated pHKM2, pHKM3 and pHKM4. All three were characterized by PCR using the degenerate primers described above. Both pHKM2 and pHKM4 gave PCR products with two sets of internal primers. pHKM3 gave a PCR product only with the FAD (SEQ ID NO: 18) and NADPH (SEQ ID NO: 19) primers suggesting that this clone likely contained a truncated *CPR* gene. All three plasmids were partially sequenced using the two Triplex primers and a third primer whose sequence was selected from the DNA sequence near the truncated end of the *CPR* gene present in pHKM1. This analysis confirmed that both pHKM2 & 4 have sequences that overlap pHKM1 and that both contained the 3' region of *CPR* gene that is missing from pHKM1. Portions of inserts from pHKM1 and pHKM4 were sequenced and a full-length *CPR* gene was identified. Based on the DNA sequence and PCR analysis, it was concluded that

pHKM1 contained the putative promoter region and 1.2 kb of sequence encoding a portion (5' end) of a *CPR* gene. pHKM4 had 1.1 kb of DNA that overlapped pHKM1 and contained the remainder (3' end) of a *CPR* gene along with a downstream untranslated region (Figure 6). Together these two plasmids contained a complete *CPRA* gene with an upstream promoter
5 region. *CPRA* is 4206 nucleotides in length (SEQ ID NO: 81) and includes a regulatory region and a protein coding region (defined by nucleotides 1006-3042) which is 2037 base pairs in length and codes for a putative protein of 679 amino acids (SEQ ID NO: 83) (Figures 13 and 14). In Figure 13, the asterisks denote conserved nucleotides between *CPRA* and *CPRB*, bold denotes protein coding nucleotides, and the start and stop codons are underlined. The *CPRA*
10 protein, when analyzed by the protein alignment program of the GeneWorks™ software package (Oxford Molecular Group, Campbell, CA), showed extensive homology to *CPR* proteins from *C. tropicalis* 750 and *C. maltosa*.

2) Cloning of the *CPRB* Allele

15 To clone the second *CPRB* allele, the third genomic library, prepared by Henkel, was screened using DNA fragments from pHKM1 and pHKM4 as probes. Five clones were obtained and these were sequenced with the three internal primers used to sequence *CPRA*. These primers were designated PRK1.F3 (SEQ ID NO: 20), PRK1.F5 (SEQ ID NO: 21) and PRK4.R20 (SEQ ID NO: 22) (Table 4). and the two outside primers (M13 -20 and T3
20 [Stratagene]) for the polylinker region present in the pBK-CMV cloning vector. Sequence analysis suggested that four of these clones, designated pHKM5 to 8, contained inserts which were identical to the *CPRA* allele isolated earlier. All four seemed to contain a full length *CPR* gene. The fifth clone was very similar to the *CPRA* allele, especially in the open reading frame region where the identity was very high. However, there were significant differences in the 5'
25 and 3' untranslated regions. This suggested that the fifth clone was the allele to *CPRA*. The plasmid was designated pHKM9 (Figure 7) and a 4.14 kb region of this plasmid was sequenced and the analysis of this sequence confirmed the presence of the *CPRB* allele (SEQ ID NO: 82), which includes a regulatory region and a protein coding region (defined by nucleotides 1033-3069) (Figure 13). The amino acid sequence of the *CPRB* protein is set forth in SEQ ID NO: 84
30 (Figure 14).

B. Cloning of *C. tropicalis* 20336 (*CYP*) Genes**1) Cloning of *CYP52A2A*, *CYP52A3A* & *3B* and *CYP52A5A* & *5B***

Clones carrying *CYP52A2A*, *A3A*, *A3B*, *A5A* and *A5B* genes were

isolated from the first and second Clontech genomic libraries using an oligonucleotide probe
5 (HemeB1, SEQ ID NO: 27) whose sequence was based upon the amino acid sequence for the
highly conserved heme binding region present throughout the *CYP52* family. The first and
second libraries were converted to the plasmid form and screened by colony hybridizations
using the HemeB1 probe (SEQ ID NO: 27) (Table 4). Several potential clones were isolated and
the plasmid DNA was isolated from these clones and sequenced using the HemeB1
10 oligonucleotide (SEQ ID NO: 27) as a primer. This approach succeeded in identifying five
CYP52 genes. Three of the *CYP* genes appeared unique, while the remaining two were classified
as alleles. Based upon an arbitrary choice of homology to *CYP52* genes from *Candida maltosa*,
these five genes and corresponding plasmids were designated *CYP52A2A* (pPA15 [Figure 26]),
CYP52A3A (pPA57 [Figure 29]), *CYP52A3B* (pPA62 [Figure 30]), *CYP52A5A* (pPAL3 [Figure
15 31]) and *CYP52A5B* (pPA5 [Figure 32]). The complete DNA sequence including regulatory and
protein coding regions of these five genes was obtained and confirmed that all five were *CYP52*
genes (Figure 15). In Figure 15, the asterisks denote conserved nucleotides among the *CYP*
genes. Bold indicates the protein coding nucleotides of the *CYP* genes, and the start and stop
codons are underlined. The *CYP52A2A* gene as represented by SEQ ID NO: 86 has a protein
20 coding region defined by nucleotides 1199-2767 and the encoded protein has an amino acid
sequence as set forth in SEQ ID NO: 96. The *CYP52A3A* gene as represented by SEQ ID NO:
88 has a protein encoding region defined by nucleotides 1126-2748 and the encoded protein has
an amino acid sequence as set forth in SEQ ID NO: 98. The *CYP52A3B* gene as represented by
SEQ ID NO: 89 has a protein coding defined by nucleotides 913-2535 and the encoded protein
25 has an amino acid sequence as set forth in SEQ ID NO: 99. The *CYP52A5A* gene as represented
by SEQ ID NO: 90 has a protein coding region defined by nucleotides 1103-2656 and the
encoded protein has an amino acid sequence as set forth in SEQ ID NO: 100. The *CYP52A5B*
gene as represented by SEQ ID NO: 91 has a protein coding region defined by nucleotides 1142-
2695 and the encoded protein has an amino acid sequence as set forth in SEQ ID NO: 101.

30

2) Cloning of *CYP52A1A* and *CYP52A8A*

CYP52A1A and *CYP52A8A* genes were isolated from the third genomic library using PCR fragments as probes. The PCR fragment probe for *CYP52A1* was generated after PCR amplification of 20336 genomic DNA with oligonucleotide primers that were designed to amplify a region from the Helix I region to the HR2 region using all available *CYP52* genes from National Center for Biotechnology Information. Degenerate forward primers UCup1 (SEQ ID NO: 23) and UCup2 (SEQ ID NO: 24) were designed based upon an amino acid sequence (-RDTTAG-) from the Helix I region (Table 4). Degenerate primers UCdown1 (SEQ ID NO: 25) and UCdown2 (SEQ ID NO: 26) were designed based upon an amino acid sequence (-GQQFAL-) from the HR2 region (Table 4). For the reverse primers, the DNA sequence represents the reverse complement of the corresponding amino acid sequence. These primers were used in pairwise combinations in a PCR reaction with Stoffel *Taq* DNA polymerase (Perkin-Elmer Cetus, Foster City, CA) according to the manufacturer's recommended procedure. A PCR product of approximately 450 bp was obtained. This product was purified from agarose gel using Gene-clean™ (Bio 101, LaJolla, CA) and ligated to the pTAG™ vector (Figure 17) (R&D systems, Minneapolis, MN) according to the recommendations of the manufacturer. No treatment was necessary to clone into pTAG because it employs the use of the TA cloning technique. Plasmids from several transformants were isolated and their inserts were characterized. One plasmid contained the PCR clone intact. The DNA sequence of the PCR fragment (designated 44CYP3, SEQ ID NO: 107) shared homology with the DNA sequences for the *CYP52A1* gene of *C. maltosa* and the *CYP52A3* gene of *C. tropicalis* 750. This fragment was used as a probe in isolating the *C. tropicalis* 20336 *CYP52A1* homolog. The third genomic library was screened using the 44CYP3 PCR probe (SEQ ID NO: 107) and a clone (pHKM11) that contained a full-length *CYP52* gene was obtained (Figure 8). The clone contained a gene having regulatory and protein coding regions. An open reading frame of 1572 nucleotides encoded a *CYP52* protein of 523 amino acids (Figures 15 and 16). This *CYP52* gene was designated *CYP52A1A* (SEQ ID NO: 85) since its putative amino acid sequence (SEQ ID NO: 95) was most similar to the *CYP52A1* protein of *C. maltosa*. The protein coding region of the *CYP52A1A* gene is defined by nucleotides 1177-2748 of SEQ ID NO: 85.

A similar approach was taken to clone *CYP52A8A*. A PCR fragment probe for *CYP52A8* was generated using primers for highly conserved sequences of *CYP52A3*, *CYP52A2*

and *CYP52A5* genes of *C. tropicalis* 750. The reverse primer (primer 2,3,5,M) (SEQ ID NO: 29) was designed based on the highly conserved heme binding region (Table 4). The design of the forward primer (primer 2,3,5,P) (SEQ ID NO: 28) was based upon a sequence conserved near the N-terminus of the *CYP52A3*, *CYP52A2* and *CYP52A5* genes from *C. tropicalis* 750 (Table 4). Amplification of 20336 genomic DNA with these two primers gave a mixed PCR product. One amplified PCR fragment was 1006 bp long (designated DCA1002). The DNA sequence for this fragment was determined and was found to have 85% identity to the DNA sequence for the *CYP52D4* gene of *C. tropicalis* 750. When this PCR product was used to screen the third genomic library one clone (pHKM12) was identified that contained a full-length *CYP52* gene along with 5' and 3' flanking sequences (Figure 9). The *CYP52* gene included regulatory and protein coding regions with an open reading frame of 1539 nucleotides long which encoded a putative *CYP52* protein of 512 amino acids (Figures 15 and 16). This gene was designated as *CYP52A8A* (SEQ ID NO: 92) since its amino acid sequence (SEQ ID NO: 102) was most similar to the *CYP52A8* protein of *C. maltosa*. The protein coding region of the *CYP52A8A* gene is defined by nucleotides 464-2002 of SEQ ID NO: 92. The amino acid sequence of the *CYP52A8A* protein is set forth in SEQ ID NO: 102.

3) Cloning of *CYP52D4A*

The screening of the second genomic library with the HemeB1 (SEQ ID NO: 27) primer (Table 4) yielded a clone carrying a plasmid (pPA18) that contained a truncated gene having homology with the *CYP52D4* gene of *C. maltosa* (Figure 33). A 1.3 to 1.5-kb *EcoRI*-*SstI* fragment from pPA18 containing part of the truncated *CYP* gene was isolated and used as a probe to screen the third genomic library for a full length *CYP52* gene. One clone (pHKM13) was isolated and found to contain a full-length *CYP* gene with extensive 5' and 3' flanking sequences (Figure 10). This gene has been designated as *CYP52D4A* (SEQ ID NO: 94) and the complete DNA including regulatory and protein coding regions (coding region defined by nucleotides 767-2266) and putative amino acid sequence (SEQ ID NO: 104) of this gene is shown in Figures 15 and 16. *CYP52D4A* (SEQ ID NO: 94) shares the greatest homology with the *CYP52D4* gene of *C. maltosa*.

30

4) Cloning of *CYP52A2B* and *CYP52A8B*

A mixed probe containing *CYP52A1A*, *A2A*, *A3A*, *D4A*, *A5A* and *A8A* genes was used to screen the third genomic library and several putative positive clones were identified. Seven of these were sequenced with the degenerate primers Cyp52a (SEQ ID NO: 32), Cyp52b (SEQ ID NO: 33), Cyp52c (SEQ ID NO: 34) and Cyp52d (SEQ ID NO: 35) shown in Table 4. These primers were designed from highly conserved regions of the four *CYP52* subfamilies, namely *CYP52A*, *B*, *C* & *D*. Sequences from two clones, pHKM14 and pHKM15 (Figures 11 and 12), shared considerable homology with DNA sequence of the *C. tropicalis* 20336 *CYP52A2* and *CYP52A8* genes, respectively. The complete DNA (SEQ ID NO: 87) including regulatory and protein coding regions (coding region defined by nucleotides 1072-2640) and putative amino acid sequence (SEQ ID NO: 97) of the *CYP52* gene present in pHKM14 suggested that it is *CYP52A2B* (Figures 15 and 16). The complete DNA (SEQ ID NO: 93) including regulatory and protein coding regions (coding region defined by nucleotides 1017-2555) and putative amino acid sequence (SEQ ID NO: 103) of the *CYP52* gene present in pHKM15 suggested that it is *CYP52A8B* (Figures 15 and 16).

EXAMPLE 14

Identification of *CYP* and *CPR* Genes Induced by Selected Fatty Acid and Alkane Substrates

Genes whose transcription is turned on by the presence of selected fatty acid or alkane substrates have been identified using the QC-RT-PCR assay. This assay was used to measure (*CYP*) and (*CPR*) gene expression in fermentor grown cultures *C. tropicalis* ATCC 20962. This method involves the isolation of total cellular RNA from cultures of *C. tropicalis* and the quantification of a specific mRNA within that sample through the design and use of sequence specific QC-RT-PCR primers and an RNA competitor. Quantification is achieved through the use of known concentrations of highly homologous competitor RNA in the QC-RT-PCR reactions. The resulting QC-RT-PCR amplified cDNA's are separated and quantitated through the use of ion pairing reverse phase HPLC. This assay was used to characterize the expression of *CYP52* genes of *C. tropicalis* ATCC 20962 in response to various fatty acid and alkane substrates. Genes which were induced were identified by the calculation of their mRNA concentration at various times before and after induction. Figure 18 provides an example of

how the concentration of mRNA for *CYP52A5* can be calculated using the QC-RT-PCR assay. The log ratio of unknown (U) to competitor product (C) is plotted versus the concentration of competitor RNA present in the QC-RT-PCR reactions. The concentration of competitor which results in a log ratio of U/C of zero, represents the point where the unknown messenger RNA concentration is equal to the concentration of the competitor. Figure 18 allows for the calculation of the amount of *CYP52A5* message present in 100 ng of total RNA isolated from cell samples taken at 0, 1, and 2 hours after the addition of Emersol® 267 in a fermentor run. From this analysis, it is possible to determine the concentration of the *CYP52A5* mRNA present in 100 ng of total cellular RNA. In the plot contained in Figure 18 it takes 0.46 pg of competitor to equal the number of mRNA's of *CYP52A5* in 100 ng of RNA isolated from cells just prior (time 0) to the addition of the substrate, Emersol® 267. In cell samples taken at one and two hours after the addition of Emersol® 267 it takes 5.5 and 8.5 pg of competitor RNA, respectively. This result demonstrates that *CYP52A5* (SEQ ID NOS: 90 and 91) is induced more than 18 fold within two hours after the addition of Emersol® 267. This type of analysis was used to demonstrate that *CYP52A5* (SEQ ID NO: 90 and 91) is induced by Emersol® 267. Figure 19 shows the relative amounts of *CYP52A5* (SEQ ID NOS: 90 and 91) expression in fermentor runs with and without Emersol® 267 as a substrate. The differences in the *CYP52A5* (SEQ. ID NOS: 90 and 91) expression patterns are due to the addition of Emersol® 267 to the fermentation medium.

This analysis clearly demonstrates that expression of *CYP52A5* (SEQ ID NOS: 90 and 91) in *C. tropicalis* 20962 is inducible by the addition of Emersol® 267 to the growth medium. This analysis was performed to characterize the expression of *CYP52A2A* (SEQ ID NO: 86), *CYP52A3AB* (SEQ ID NOS: 88 and 89), *CYP52A8A* (SEQ ID NO: 92), *CYP52A1A* (SEQ ID NO: 85), *CYP52D4A* (SEQ ID NO: 94) and *CPRB* (SEQ ID NO: 82) in response to the presence of Emersol® 267 in the fermentation medium (Figure 20). The results of these analysis' indicate, that like the *CYP52A5* gene (SEQ ID NOS: 90 and 91) of *C. tropicalis* 20962, the *CYP52A2A* gene (SEQ ID NO: 86) is inducible by Emersol® 267. A small induction is observed for *CYP52A1A* (SEQ ID NO: 85) and *CYP52A8A* (SEQ ID NO: 92). In contrast, any induction for *CYP52D4A* (SEQ ID NO: 94), *CYP52A3A* (SEQ ID NO: 88), *CYP52A3B* (SEQ ID NO: 89) is below the level of detection of the assay. *CPRB* (SEQ ID NO: 82) is moderately induced by Emersol® 267, four to five fold. The results of these analysis are summarized in

Figure 20. Figure 34 provides an example of selective induction of *CYP52A* genes. When pure fatty acid or alkanes are spiked into a fermentor containing *C. tropicalis* 20962 or a derivative thereof, the transcriptional activation of *CYP52A* genes was detected using the QC-RT-PCR assay. Figure 34 shows that pure oleic acid (C18:1) strongly induces *CYP52A2A* (SEQ ID NO: 86) while inducing *CYP52A5* (SEQ ID NOS: 90 and 91). In the same fermentor addition of pure alkane (tridecane) shows strong induction of both *CYP52A2A* (SEQ ID NO: 86) and *CYP52A1A* (SEQ ID NO: 85). However, tridecane did not induce *CYP52A5* (SEQ ID NOS: 90 and 91). In a separate fermentation using ATCC 20962, containing pure octadecane as the substrate, induction of *CYP52A2A*, *CYP52A5A* and *CYP52A1A* is detected (see Figure 36). The foregoing demonstrates selective induction of particular *CYP* genes by specific substrates, thus providing techniques for selective metabolic engineering of cell strains. For example, if tridecane modification is desired, organisms engineered for high levels of *CYP52A2A* (SEQ ID NO: 86) and *CYP52A1A* (SEQ ID NO: 85) activity are indicated. If oleic acid modification is desired, organisms engineered for high levels of *CYP52A2A* (SEQ ID NO: 86) activity are indicated.

EXAMPLE 15

Integration of Selected *CYP* and *CPR* Genes into the Genome of *Candida tropicalis*

In order to integrate selected genes into the chromosome of *C. tropicalis* 20336 or its descendants, there has to be a target DNA sequence, which may or may not be an intact gene, into which the genes can be inserted. There must also be a method to select for the integration event. In some cases the target DNA sequence and the selectable marker are the same and, if so, then there must also be a method to regain use of the target gene as a selectable marker following the integration event. In *C. tropicalis* and its descendants, one gene which fits these criteria is *URA3A*, encoding orotidine-5'-phosphate decarboxylase. Using it as a target for integration, *ura* variants of *C. tropicalis* can be transformed in such a way as to regenerate a *URA*⁺ genotype via homologous recombination (Figure 21). Depending upon the design of the integration vector, one or more genes can be integrated into the genome at the same time. Using a split *URA3A* gene oriented as shown in Figure 22, homologous integration would yield at least one copy of the gene(s) of interest which are inserted between the split portions of the *URA3A* gene. Moreover, because of the high sequence similarity between *URA3A* and *URA3B* genes, integration of the

Figure 20. Figure 34 provides an example of selective induction of *CYP52A* genes. When pure fatty acid or alkanes are spiked into a fermentor containing *C. tropicalis* 20962 or a derivative thereof, the transcriptional activation of *CYP52A* genes was detected using the QC-RT-PCR assay. Figure 34 shows that pure oleic acid (C18:1) strongly induces *CYP52A2A* (SEQ ID NO: 86) while inducing *CYP52A5* (SEQ ID NOS: 90 and 91). In the same fermentor addition of pure alkane (tridecane) shows strong induction of both *CYP52A2A* (SEQ ID NO: 86) and *CYP52A1A* (SEQ ID NO: 85). However, tridecane did not induce *CYP52A5* (SEQ ID NOS: 90 and 91). In a separate fermentation using ATCC 20962, containing pure octadecane as the substrate, induction of *CYP52A2A*, *CYP52A5A* and *CYP52A1A* is detected (see Figure 36). The foregoing demonstrates selective induction of particular *CYP* genes by specific substrates, thus providing techniques for selective metabolic engineering of cell strains. For example, if tridecane modification is desired, organisms engineered for high levels of *CYP52A2A* (SEQ ID NO: 86) and *CYP52A1A* (SEQ ID NO: 85) activity are indicated. If oleic acid modification is desired, organisms engineered for high levels of *CYP52A2A* (SEQ ID NO: 86) activity are indicated.

15

EXAMPLE 15

Integration of Selected *CYP* and *CPR* Genes into the Genome of *Candida tropicalis*

In order to integrate selected genes into the chromosome of *C. tropicalis* 20336 or its descendants, there has to be a target DNA sequence, which may or may not be an intact gene, into which the genes can be inserted. There must also be a method to select for the integration event. In some cases the target DNA sequence and the selectable marker are the same and, if so, then there must also be a method to regain use of the target gene as a selectable marker following the integration event. In *C. tropicalis* and its descendants, one gene which fits these criteria is *URA3A*, encoding orotidine-5'-phosphate decarboxylase. Using it as a target for integration, *ura* variants of *C. tropicalis* can be transformed in such a way as to regenerate a *URA*⁺ genotype via homologous recombination (Figure 21). Depending upon the design of the integration vector, one or more genes can be integrated into the genome at the same time. Using a split *URA3A* gene oriented as shown in Figure 22, homologous integration would yield at least one copy of the gene(s) of interest which are inserted between the split portions of the *URA3A* gene. Moreover, because of the high sequence similarity between *URA3A* and *URA3B* genes, integration of the

25
30

construct can occur at both the *URA3A* and *URA3B* loci. Subsequently, an oligonucleotide designed with a deletion in a portion of the *URA* gene based on the identical sequence across both the *URA3A* and *URA3B* genes, can be utilized to yield *C. tropicalis* transformants which are once again *ura⁻* but which still carry one or more newly integrated genes of choice (Figure 5 21). *ura⁻* variants of *C. tropicalis* can also be isolated via other methods such as classical mutagenesis or by spontaneous mutation. Using well established protocols, selection of *ura⁻* strains can be facilitated by the use of 5-fluoroorotic acid (5-FOA) as described, e.g., in Boeke et al., *Mol. Gen. Genet.* 197:345-346, (1984), incorporated herein by reference. The utility of this approach for the manipulation of *C. tropicalis* has been well documented as described, e.g., in 10 Picataggio et al., *Mol. and Cell. Biol.* 11:4333-4339 (1991); Rohrer et al., *Appl. Microbiol. Biotechnol.* 36:650-654 (1992); Picataggio et al., *Bio/Technology* 10:894-898 (1992); U.S. Patent No. 5,648,247; U.S. Patent No. 5,620,878; U.S. Patent No. 5,204,252; U.S. Patent No. 5,254,466, all of which are incorporated herein by reference.

15 A. Construction of a URA Integration Vector, pURAI_n.

Primers were designed and synthesized based on the 1712 bp sequence of the *URA3A* gene of *C. tropicalis* 20336 (see Figure 23). The nucleotide sequence of the *URA3A* gene of *C. tropicalis* 20336 is set forth in SEQ ID NO: 105 and the amino acid sequence of the encoded protein is set forth in SEQ ID NO: 106. *URA3A* Primer Set #1a (SEQ ID NO: 9) and 20 #1b (SEQ ID NO: 10) (Table 4) was used in PCR with *C. tropicalis* 20336 genomic DNA to amplify *URA3A* sequences between nucleotide 733 and 1688 as shown in Figure 23. The primers are designed to introduce unique 5' *Asc*I and 3' *Pac*I restriction sites into the resulting amplified *URA3A* fragment. *Asc*I and *Pac*I sites were chosen because these sites are not present within *CYP* or *CPR* genes identified to date. *URA3A* Primer Set #2 was used in PCR with *C.* 25 *tropicalis* 20336 genomic DNA as a template, to amplify *URA3A* sequences between nucleotide 9 and 758 as shown in Figure 23. *URA3A* Primer set #2a (SEQ ID NO: 11) and #2b (SEQ ID NO: 12) (Table 4) was designed to introduce unique 5' *Pac*I and 3' *Pme*I restriction sites into the resulting amplified *URA3A* fragment. The *Pme*I site is also not present within *CYP* and *CPR* genes identified to date. PCR fragments of the *URA3A* gene were purified, restricted with *Asc*I, 30 *Pac*I and *Pme*I restriction enzymes and ligated to a gel purified, QiaexII cleaned *Asc*I-*Pme*I digest of plasmid pNEB193 (Figure 25) purchased from New England Biolabs (Beverly, MA).

The ligation was performed with an equimolar number of DNA termini at 16 °C for 16 hr using T4 DNA ligase (New England Biolabs). Ligations were transformed into *E. coli* XL1-Blue cells (Stratagene, LaJolla, CA) according to manufacturers recommendations. White colonies were isolated, grown, plasmid DNA isolated and digested with *AscI-PmeI* to confirm insertion of the modified *URA3A* into pNEB193. The resulting base integration vector was named pURAin (Figure 24).

B. Amplification of *CYP52A2A*, *CYP52A3A*, *CYP52A5A* and *CPRB* from *C. tropicalis* 20336 Genomic DNA

The genes encoding *CYP52A2A*, (SEQ ID NO: 86) and *CYP52A3A* (SEQ ID NO: 88) from *C. tropicalis* 20336 were amplified from genomic clones (pPA15 and pPA57, respectively) (Figures 26 and 29) via PCR using primers (Primer *CYP 2A#1*, SEQ ID NO: 1 and Primer *CYP 2A#2*, SEQ ID NO: 2 for *CYP52A2A*) (Primer *CYP 3A#1*, SEQ ID NO: 3 and Primer *CYP 3A#2*, SEQ ID NO: 4 for *CYP52A3A*) to introduce *PacI* cloning sites. These PCR primers were designed based upon the DNA sequence determined for *CYP52A2A* (SEQ ID NO: 86) (Figure 15). The *AmpliTaq* Gold PCR kit (Perkin Elmer Cetus, Foster City, CA) was used according to manufacturers specifications. The *CYP52A2A* PCR amplification product was 2,230 base pairs in length, yielding 496 bp of DNA upstream of the *CYP52A2A* start codon and 168 bp downstream of the stop codon for the *CYP52A2A* ORF. The *CYP52A3A* PCR amplification product was 2154 base pairs in length, yielding 437bp of DNA upstream of the *CYP52A3A* start codon and 97bp downstream of the stop codon for the *CYP52A3A* ORF. The *CYP52A3A* PCR amplification product was 2154 base pairs in length, yielding 437bp of DNA upstream of the *CYP52A3A* start codon and 97bp downstream of the stop codon for the *CYP52A3A* ORF.

The gene encoding *CYP52A5A* (SEQ ID NO: 90) from *C. tropicalis* 20336 was amplified from genomic DNA via PCR using primers (Primer *CYP 5A#1*, SEQ ID NO: 5 and Primer *CYP 5A#2*, SEQ ID NO: 6) to introduce *PacI* cloning sites. These PCR primers were designed based upon the DNA sequence determined for *CYP52A5A* (SEQ ID NO: 90). The *Expand Hi-Fi Taq* PCR kit (Boehringer Mannheim, Indianapolis, IN) was used according to manufacturers specifications. The *CYP52A5A* PCR amplification product was 3,298 base pairs in length.

The gene encoding *CPRB* (SEQ ID NO: 82) from *C. tropicalis* 20336 was amplified from genomic DNA via PCR using primers (*CPR B#1*, SEQ ID NO: 7 and *CPR B#2*, SEQ ID NO: 8) based upon the DNA sequence determined for *CPRB* (SEQ ID NO: 82) (Figure 13). These primers were designed to introduce unique *PacI* cloning sites. The Expand Hi-Fi
5 *Taq* PCR kit (Boehringer Mannheim, Indianapolis, IN) was used according to manufacturers specifications. The *CPRB* PCR product was 3266 bp in length, yielding 747 bp pf DNA upstream of the *CPRB* start codon and 493 bp downstream of the stop codon for the *CPRB* ORF. The resulting PCR products were isolated via agarose gel electrophoresis, purified using QiaexII and digested with *PacI*. The PCR fragments were purified, desalted and concentrated using a
10 Microcon 100 (Amicon, Beverly, MA).

The above described amplification procedures are applicable to the other genes listed in Table 5 using the respectively indicated primers.

C. Cloning of *CYP* and *CPR* Genes into pURain.

15 The next step was to clone the selected *CYP* and *CPR* genes into the pURain integration vector. In a preferred aspect of the present invention, no foreign DNA other than that specifically provided by synthetic restriction site sequences are incorporated into the DNA which was cloned into the genome of *C. tropicalis*, i.e., with the exception of restriction site DNA only native *C. tropicalis* DNA sequences are incorporated into the genome. pURain was digested
20 with *PacI*, Qiaex II cleaned, and dephosphorylated with Shrimp Alkaline Phosphatase (SAP) (United States Biochemical, Cleveland, OH) according to the manufacturer's recommendations. Approximately 500 ng of *PacI* linearized pURain was dephosphorylated for 1 hr at 37°C using SAP at a concentration of 0.2 Units of enzyme per 1 pmol of DNA termini. The reaction was stopped by heat inactivation at 65°C for 20 min.

25 The *CYP52A2A* *PacI* fragment derived using the primer shown in Table 4 was ligated to plasmid pURain which had also been digested with *PacI*. *PacI* digested pURain was dephosphorylated, and ligated to the *CYP52A2A* ULTMA PCR product as described previously. The ligation mixture was transformed into *E. coli* XL1 Blue MRF' (Stratagene) and 2 resistant colonies were selected and screened for correct constructs which should contain vector sequence,
30 the inverted *URA3A* gene, and the amplified *CYP52A2A* gene (SEQ ID NO: 86) of 20336. *AscI*-*PmeI* digestion identified one of the two constructs, plasmid pURA2in, as being correct (Figure

27). This plasmid was sequenced and compared to *CYP52A2A* (SEQ ID NO: 86) to confirm that PCR did not introduce DNA base changes that would result in an amino acid change.

Prior to its use, the *CPRB* *PacI* fragment derived using the primers shown in Table 4 was sequenced and compared to *CPRB* (SEQ ID NO: 82) to confirm that PCR did not
5 introduce DNA base pair changes that would result in an amino acid change. Following confirmation, *CPRB* (SEQ ID NO: 82) was ligated to plasmid pURAI_n which had also been digested with *PacI*. *PacI* digested pURAI_n was dephosphorylated, and ligated to the *CPR* Expand Hi-Fi PCR product as described previously. The ligation mixture was transformed into
10 *E. coli* XL1 Blue MRF' (Stratagene) and several resistant colonies were selected and screened for correct constructs which should contain vector sequence, the inverted *URA3A* gene, and the amplified *CPRB* gene (SEQ ID NO: 82) of 20336. *AscI*-*PmeI* digestion confirmed a successful construct, pURAREDBin.

In a manner similar to the above, each of the other *CYP* and *CPR* genes disclosed herein are cloned into pURAI_n. *PacI* fragments of these genes, whose sequences are given in
15 Figures 13 and 15, are derivable by methods known to those skilled in the art.

1) Construction of Vectors Used to Generate HDC 20 and HDC 23

A previously constructed integration vector containing *CPRB* (SEQ ID NO: 82), pURAREDBin, was chosen as the starting vector. This vector was partially digested with *PacI*
20 and the linearized fragment was gel-isolated. The active *PacI* was destroyed by treatment with T4 DNA polymerase and the vector was re-ligated. Subsequent isolation and complete digestion of this new plasmid yielded a vector now containing only one active *PacI* site. This fragment was gel-isolated, dephosphorylated and ligated to the *CYP52A2A* *PacI* fragment. Vectors that contain the *CYP52A2A* (SEQ ID NO: 86) and *CPRB* (SEQ ID NO: 82) genes oriented in the
25 same direction, pURAI_n *CPR* 2A S, as well as opposite directions (5' ends connected), pURAI_n *CPR* 2A O, were generated.

D. Confirmation of *CYP* Integration (Figure 21 for Integration Scheme) into the Genome of *C. tropicalis*

30 Based on the construct, pURA2in, used to transform H5343 *ura'*, a scheme to detect integration was devised. Genomic DNA from transformants was digested with *Dra* III

and *Spe I* which are enzymes that cut within the *URA3A*, and *URA3B* genes but not within the integrated *CYP52A2A* gene. Digestion of genomic DNA where an integration had occurred at the *URA3A* or *URA3B* loci would be expected to result in a 3.5 kb or a 3.3 kb fragment, respectively (Figure 28). Moreover, digestion of the same genomic DNA with *PacI* would yield a 2.2 kb fragment characteristic for the integrated *CYP52A2A* gene (Figure 28). Southern hybridizations of these digests with fragments of the *CYP52A2A* gene were used to screen for these integration events. Intensity of the band signal from the Southern using *PacI* digestion was used as a measure of the number of integration events, ((i.e. the more copies of the *CYP52A2A* gene (SEQ ID NO: 86) which are present, the stronger the hybridization signal)).

5
10
15
20
C. tropicalis H5343 transformed *URA* prototrophs were grown at 30°C, 170 rpm, in 10 ml SC-uracil media for preparation of genomic DNA. Genomic DNA was isolated by the method described previously. Genomic DNA was digested with *SpeI* and *DraIII*. A 0.95% agarose gel was used to prepare a Southern hybridization blot. The DNA from the gel was transferred to a MagnaCharge nylon filter membrane (MSI Technologies, Westboro, MA) according to the alkaline transfer method of Sambrook et al., *supra*. For the Southern hybridization, a 2.2 kb *CYP52A2A* DNA fragment was used as a hybridization probe. 300 ng of *CYP52A2A* DNA was labeled using a ECL Direct labeling and detection system (Amersham) and the Southern was processed according to the ECL kit specifications. The blot was processed in a volume of 30 ml of hybridization fluid corresponding to 0.125 ml/cm². Following a prehybridization at 42°C for 1 hr, 300 ng of *CYP52A2A* probe was added and the hybridization continued for 16 hr at 42°C. Following hybridization, the blots were washed two times for 20 min each at 42 °C in primary wash containing urea. Two 5 min secondary washes at RT were conducted, followed by detection according to directions. The blots were exposed for 16 hours (hr) as recommended.

25
Integration was confirmed by the detection of a *SpeI*-*DraIII* 3.5 kb fragment from the genomic DNA of the transformants but not with the *C. tropicalis* 20336 control. Subsequently, a *PacI* digestion of the genomic DNA of the positive transformants, followed by a Southern hybridization using an *CYP52A2A* gene probe, confirmed integration by the detection of a 2.2 kb fragment. The resulting *CYP52A2A* integrated strain was named HDC1 (see Table 1).

and *Spe I* which are enzymes that cut within the *URA3A*, and *URA3B* genes but not within the integrated *CYP52A2A* gene. Digestion of genomic DNA where an integration had occurred at the *URA3A* or *URA3B* loci would be expected to result in a 3.5 kb or a 3.3 kb fragment, respectively (Figure 28). Moreover, digestion of the same genomic DNA with *PacI* would yield a 2.2 kb fragment characteristic for the integrated *CYP52A2A* gene (Figure 28). Southern hybridizations of these digests with fragments of the *CYP52A2A* gene were used to screen for these integration events. Intensity of the band signal from the Southern using *PacI* digestion was used as a measure of the number of integration events, ((i.e. the more copies of the *CYP52A2A* gene (SEQ ID NO: 86) which are present, the stronger the hybridization signal)).

10 *C. tropicalis* H5343 transformed *URA* prototrophs were grown at 30°C, 170 rpm, in 10 ml SC-uracil media for preparation of genomic DNA. Genomic DNA was isolated by the method described previously. Genomic DNA was digested with *SpeI* and *DraIII*. A 0.95% agarose gel was used to prepare a Southern hybridization blot. The DNA from the gel was transferred to a MagnaCharge nylon filter membrane (MSI Technologies, Westboro, MA) according to the alkaline transfer method of Sambrook et al., *supra*. For the Southern hybridization, a 2.2 kb *CYP52A2A* DNA fragment was used as a hybridization probe. 300 ng of *CYP52A2A* DNA was labeled using a ECL Direct labeling and detection system (Amersham) and the Southern was processed according to the ECL kit specifications. The blot was processed in a volume of 30 ml of hybridization fluid corresponding to 0.125 ml/cm². Following a prehybridization at 42°C for 1 hr, 300 ng of *CYP52A2A* probe was added and the hybridization continued for 16 hr at 42°C. Following hybridization, the blots were washed two times for 20 min each at 42 °C in primary wash containing urea. Two 5 min secondary washes at RT were conducted, followed by detection according to directions. The blots were exposed for 16 hours (hr) as recommended.

25 Integration was confirmed by the detection of a *SpeI-DraIII* 3.5 kb fragment from the genomic DNA of the transformants but not with the *C. tropicalis* 20336 control. Subsequently, a *PacI* digestion of the genomic DNA of the positive transformants, followed by a Southern hybridization using an *CYP52A2A* gene probe, confirmed integration by the detection of a 2.2 kb fragment. The resulting *CYP52A2A* integrated strain was named HDC1 (see Table 1).

In a manner similar to the above, each of the genes contained in the *PacI* fragments which are described in Section 3c above were confirmed for integration into the genome of *C. tropicalis*.

Transformants generated by transformation with the vectors, pURAI in *CPR* 2A S or pURAI in *CPR* 2A O, were analyzed by Southern hybridization for integration of both the *CYP52A2A* (SEQ ID NO: 86) and *CPRB* (SEQ ID NO: 82) genes tandemly. Three strains were generated in which the *CYP52A2A* (SEQ ID NO: 86) and *CPRB* (SEQ ID NO: 82) genes integrated are in the opposite orientation (HDC 20-1, HDC 20-2 and HDC 20-3) and three were generated with the *CYP52A2A* (SEQ ID NO: 86) and *CPRB* (SEQ ID NO: 82) genes integrated in the same orientation (HDC 23-1, HDC 23-2 and HDC 23-3), Table 1.

E. Confirmation of *CPRB* Integration into H5343 *ura*

Seven transformants were screened by colony PCR using *CPRB* primer #2 (SEQ ID NO: 8) and a *URA3A*- specific primer. In five of the transformants, successful integration was detected by the presence of a 3899 bp PCR product. This 3899 bp PCR product represents the *CPRB* gene adjacent to the *URA3A* gene in the genome of H5343 thereby confirming integration. The resulting *CPRB* integrated strains were named HDC10-1 and HDC10-2 (see Table 1).

F. Strain Evaluation.

As determined by quantitative PCR, when compared to parent H5343, HDC10-1 contained three additional copies of the reductase gene and HDC10-2 contained four additional copies of the reductase gene. Evaluations of HDC20-1, HDC20-2 and HDC20-3 based on Southern hybridization data indicates that HDC20-1 contained multiple integrations, i.e., 2 to 3 times that of HDC20-2 or HDC20-3. Evaluations of HDC23-1, HDC23-2, and HDC23-3 based on Southern hybridization data indicates that HDC23-3 contained multiple integrations, i.e., 2 to 3 times that of HDC23-1 or HDC23-2. The data in Table 8 indicates that the integration of components of the ω -hydroxylase complex have a positive effect on the improvement of *Candida tropicalis* ATCC 20962 as a biocatalyst. The results indicate that *CYP52A5A* (SEQ ID NO: 90) is an important gene for the conversion of oleic acid to diacid. Surprisingly, tandem integrations of *CYP* and *CPR* genes oriented in the opposite direction (HDC 20 strains) seem to

be less productive than tandem integrations oriented in the same direction (HDC 23 strains),
Tables 1 and 8.

CHART

5	<u>Media Composition</u>		Magnesium Sulfate	0.98 g
			(anhydrous)	
	<u>LB Broth</u>		Agar	15 g
	Bacto Tryptone	10 g	Distilled Water	1,000 ml
10	Bacto Yeast Extract	5 g	<u>NZCYM Top Agarose</u>	
	Sodium Chloride	10 g	Bacto Casein Digest	10 g
	Distilled Water	1,000 ml	Bacto Casamino Acids	1 g
			Bacto Yeast Extract	5 g
	<u>LB Agar</u>		Sodium Chloride	5 g
15	Bacto Tryptone	10 g	Magnesium Sulfate	0.98 g
	Bacto Yeast Extract	5 g	(anhydrous)	
	Sodium Chloride	10 g	Agarose	7 g
	Agar	15 g	Distilled Water	1,000 ml
	Distilled Water	1,000 ml	<u>YEPD Broth</u>	
20	<u>LB Top Agarose</u>		Bacto Yeast Extract	10 g
	Bacto Tryptone	10 g	Bacto Peptone	20 g
	Bacto Yeast Extract	5 g	Glucose	20 g
	Sodium Chloride	10 g	Distilled Water	1,000 ml
25	Agarose	7 g	<u>YEPD Agar*</u>	
	Distilled Water	1,000 ml	Bacto Yeast Extract	10 g
			Bacto Peptone	20 g
	<u>NZCYM Broth</u>		Glucose	20 g
	Bacto Casein Digest	10 g	Agar	20 g
30	Bacto Casamino Acids	1 g	Distilled Water	1,000 ml
	Bacto Yeast Extract	5 g	<u>SC - uracil*</u>	
	Sodium Chloride	5 g	Bacto-yeast nitrogen base without amino acids	6.7g
	Magnesium Sulfate	0.98 g	Glucose	20g
	(anhydrous)		Bacto-agar	20g
35	Distilled Water	1,000 ml	Drop-out mix	2g
			Distilled water	1,000ml
	<u>NZCYM Agar</u>			
	Bacto Casein Digest	10 g		
	Bacto Casamino Acids	1 g		
40	Bacto Yeast Extract	5 g		
	Sodium Chloride	5 g		

	<u>DCA2 medium</u>	g/l
	Peptone	3.0
	Yeast Extract	6.0
	Sodium Acetate	3.0
5	Yeast Nitrogen Base (Difco)	6.7
	Glucose (anhydrous)	50.0
	Potassium Phosphate (dibasic, trihydrate)	7.2
	Potassium Phosphate (monobasic, anhydrous)	9.3

10

	<u>DCA3 medium</u>	g/l
	0.3 M Phosphate buffer containing, pH 7.5	
	Glycerol	50
	Yeast Nitrogen base (Difco)	6.7

15

<u>Drop-out mix</u>				
20	Adenine	0.5g	Alanine	2g
	Arginine	2g	Asparagine	2g
	Aspartic acid	2g	Cysteine	2g
	Glutamine	2g	Glutamic acid	2g
	Glycine	2g	Histidine	2g
	Inositol	2g	Isoleucine	2g
	Leucine	10g	Lysine	2g
	Methionine	2g	para-Aminobenzoic acid	0.2g
25	Phenylalanine	2g	Proline	2g
	Serine	2g	Threonine	2g
	Tryptophan	2g	Tyrosine	2g
	Valine	2g		

30

*See Kaiser et al., Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, USA (1994), incorporated herein by reference.

It will be understood that various modifications may be made to the embodiments and/or examples disclosed herein. Thus, the above description should not be construed as limiting, but merely as exemplifications of preferred embodiments. Those skilled in the art will envision other modifications within the scope and spirit of the claims appended hereto.

WHAT IS CLAIMED IS:

1. Isolated nucleic acid encoding a *CPRA* protein having the amino acid sequence set forth in SEQ ID NO: 83.
5
2. Isolated nucleic acid comprising a coding region defined by nucleotides 1006-3042 as set forth in SEQ ID NO: 81.
3. Isolated nucleic acid according to claim 2 comprising the nucleotide sequence
10 as set forth in SEQ ID NO: 81.
4. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 83.
5. A vector comprising a nucleotide sequence encoding *CPRA* protein including an amino acid sequence as set forth in SEQ ID NO: 83.
15
6. A vector according to claim 5 wherein the nucleotide sequence is set forth in nucleotides 1006-3042 of SEQ ID NO: 81.
20
7. A vector according to claim 5 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.
8. A host cell transfected or transformed with the nucleic acid of claim 1.
25
9. A host cell according to claim 8 wherein the host cell is a yeast cell.
10. A host cell according to claim 9 wherein the yeast cell is a *Candida sp.*
- 30 11. A host cell according to claim 10 wherein the *Candida sp.* is *Candida tropicalis*.

12. A host cell according to claim 11 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

13. A host cell according to claim 12 wherein the *Candida tropicalis* is H5343
5 ura-.

14. A method of producing a *CPRA* protein including an amino acid sequence as set forth in SEQ ID NO: 83 comprising:

- a) transforming a suitable host cell with a DNA sequence that encodes the protein
10 having the amino acid sequence as set forth in SEQ ID NO: 83; and
b) culturing the cell under conditions favoring the expression of the protein.

15. The method according to claim 14 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.
15

16. Isolated nucleic acid encoding a *CPRB* protein having the amino acid sequence set forth in SEQ ID NO: 84.

17. Isolated nucleic acid comprising a coding region defined by nucleotides 1033-
20 3069 as set forth in SEQ ID NO: 82.

18. Isolated nucleic acid according to claim 17 comprising the nucleotide sequence as set forth in SEQ ID NO: 82.

19. Isolated protein comprising an amino acid sequence as set forth in SEQ ID
25 NO: 84.

20. A vector comprising a nucleotide sequence encoding *CPRB* protein including an amino acid sequence as set forth in SEQ ID NO: 84.
30

21. A vector according to claim 20 wherein the nucleotide sequence is set forth in nucleotides 1033-3069 of SEQ ID NO: 82.

22. A vector according to claim 20 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid..

23. A host cell transfected or transformed with the nucleic acid of claim 16.

5

24. A host cell according to claim 23 wherein the host cell is a yeast cell.

25. A host cell according to claim 24 wherein the yeast cell is a *Candida sp.*

10

26. A host cell according to claim 25 wherein the *Candida sp.* is *Candida tropicalis*.

27. A host cell according to claim 26 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

15

28. A host cell according to claim 27 wherein the *Candida tropicalis* is H5343 ura-.

29. A method of producing a *CPRB* protein including an amino acid sequence as set forth in SEQ ID NO: 84 comprising:

20

- a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 84; and
- b) culturing the cell under conditions favoring the expression of the protein.

25

30. The method according to claim 29 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

31. Isolated nucleic acid encoding a *CYP52A1A* protein having the amino acid sequence set forth in SEQ ID NO: 95.

30

32. Isolated nucleic acid comprising a coding region defined by nucleotides 1177-2748 as set forth in SEQ ID NO: 85.

33. Isolated nucleic acid according to claim 32 comprising the nucleotide sequence as set forth in SEQ ID NO: 85.

5 34. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 95.

35. A vector comprising a nucleotide sequence encoding *CYP52A1A* protein including an amino acid sequence as set forth in SEQ ID NO: 95.

10 36. A vector according to claim 35 wherein the nucleotide sequence is set forth in nucleotides 1177-2748 of SEQ ID NO: 85.

15 37. A vector according to claim 35 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

38. A host cell transfected or transformed with the nucleic acid of claim 31.

39. A host cell according to claim 38 wherein the host cell is a yeast cell.

20 40. A host cell according to claim 39 wherein the yeast cell is a *Candida sp.*

41. A host cell according to claim 40 wherein the *Candida sp.* is *Candida tropicalis*.

25 42. A host cell according to claim 41 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

30 43. A host cell according to claim 42 wherein the *Candida tropicalis* is H5343 ura-.

44. A method of producing a *CYP52A1A* protein including an amino acid sequence as set forth in SEQ ID NO: 95 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 95; and

5 b) culturing the cell under conditions favoring the expression of the protein.

45. The method according to claim 44 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

10 46. Isolated nucleic acid encoding a *CYP52A2A* protein having the amino acid sequence set forth in SEQ ID NO: 96.

47. Isolated nucleic acid comprising a coding region defined by nucleotides 1199-2767 as set forth in SEQ ID NO: 86.

15

48. Isolated nucleic acid according to claim 47 comprising the nucleotide sequence as set forth in SEQ ID NO: 86.

49. Isolated protein comprising an amino acid sequence as set forth in SEQ ID
20 NO: 96.

50. A vector comprising a nucleotide sequence encoding *CYP52A2A* protein including an amino acid sequence as set forth in SEQ ID NO: 96.

25 51. A vector according to claim 50 wherein the nucleotide sequence is set forth in nucleotides 1199-2767 of SEQ ID NO: 86.

52. A vector according to claim 50 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

30

53. A host cell transfected or transformed with the nucleic acid of claim 46.

54. A host cell according to claim 53 wherein the host cell is a yeast cell.

55. A host cell according to claim 54 wherein the yeast cell is a *Candida sp.*

5 56. A host cell according to claim 55 wherein the *Candida sp.* is *Candida tropicalis*.

57. A host cell according to claim 56 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

10 58. A host cell according to claim 57 wherein the *Candida tropicalis* is H5343 ura-.

59. A method of producing a *CYP52A2A* protein including an amino acid
15 sequence as set forth in SEQ ID NO: 96 comprising:

- a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 96; and
- b) culturing the cell under conditions favoring the expression of the protein.

20 60. The method according to claim 59 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

61. Isolated nucleic acid encoding a *CYP52A2B* protein having the amino acid
sequence set forth in SEQ ID NO: 97.

25 62. Isolated nucleic acid comprising a coding region defined by nucleotides 1072-2640 as set forth in SEQ ID NO: 87.

63. Isolated nucleic acid according to claim 62 comprising the nucleotide sequence
30 as set forth in SEQ ID NO: 87.

64. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 97.

65. A vector comprising a nucleotide sequence encoding *CYP52A2B* protein including an amino acid sequence as set forth in SEQ ID NO: 97.

66. A vector according to claim 65 wherein the nucleotide sequence is set forth in nucleotides 1072-2640 of SEQ ID NO: 87.

67. A vector according to claim 65 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

68. A host cell transfected or transformed with the nucleic acid of claim 61.

69. A host cell according to claim 68 wherein the host cell is a yeast cell.

70. A host cell according to claim 69 wherein the yeast cell is a *Candida sp.*

71. A host cell according to claim 70 wherein the *Candida sp.* is *Candida tropicalis*.

72. A host cell according to claim 71 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

73. A host cell according to claim 72 wherein the *Candida tropicalis* is H5343 ura-.

74. A method of producing a *CYP52A2B* protein including an amino acid sequence as set forth in SEQ ID NO: 97 comprising:

- a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 97; and
- b) culturing the cell under conditions favoring the expression of the protein.

75. The method according to claim 74 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

5 76. Isolated nucleic acid encoding a *CYP52A3A* protein having the amino acid sequence set forth in SEQ ID NO: 98.

77. Isolated nucleic acid comprising a coding region defined by nucleotides 1126-2748 as set forth in SEQ ID NO: 88.

10 78. Isolated nucleic acid according to claim 77 comprising the nucleotide sequence as set forth in SEQ ID NO: 88.

79. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 98.
15

80. A vector comprising a nucleotide sequence encoding *CYP52A3A* protein including an amino acid sequence as set forth in SEQ ID NO: 98.

81. A vector according to claim 80 wherein the nucleotide sequence is set forth in
20 nucleotides 1126-2748 of SEQ ID NO: 88.

82. A vector according to claim 80 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

25 83. A host cell transfected or transformed with the nucleic acid of claim 76.

84. A host cell according to claim 83 wherein the host cell is a yeast cell.

85. A host cell according to claim 84 wherein the yeast cell is a *Candida sp.*
30

86. A host cell according to claim 85 wherein the *Candida sp.* is *Candida tropicalis*.

87. A host cell according to claim 86 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

5 88. A host cell according to claim 87 wherein the *Candida tropicalis* is H5343
ura-.

89. A method of producing a *CYP52A3A* protein including an amino acid sequence as set forth in SEQ ID NO: 98 comprising:

- 10 a) transforming a suitable host cell with a DNA sequence that encodes the protein
having the amino acid sequence as set forth in SEQ ID NO: 98; and
b) culturing the cell under conditions favoring the expression of the protein.

90. The method according to claim 89 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

15

91. Isolated nucleic acid encoding a *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99.

20 92. Isolated nucleic acid comprising a coding region defined by nucleotides 913-
2535 as set forth in SEQ ID NO: 89.

93. Isolated nucleic acid according to claim 92 comprising the nucleotide sequence as set forth in SEQ ID NO: 89.

25 94. Isolated protein comprising an amino acid sequence as set forth in SEQ ID
NO: 99.

95. A vector comprising a nucleotide sequence encoding *CYP52A3B* protein including an amino acid sequence as set forth in SEQ ID NO: 99.

30

96. A vector according to claim 95 wherein the nucleotide sequence is set forth in nucleotides 913-2535 of SEQ ID NO: 89.

97. A vector according to claim 95 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

98. A host cell transfected or transformed with the nucleic acid of claim 91.

5

99. A host cell according to claim 98 wherein the host cell is a yeast cell.

100. A host cell according to claim 99 wherein the yeast cell is a *Candida sp.*

10

101. A host cell according to claim 100 wherein the *Candida sp.* is *Candida tropicalis*.

102. A host cell according to claim 101 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

15

103. A host cell according to claim 102 wherein the *Candida tropicalis* is H5343 ura-.

104. A method of producing a *CYP52A3B* protein including an amino acid sequence as set forth in SEQ ID NO: 99 comprising:

20

- a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 99; and
- b) culturing the cell under conditions favoring the expression of the protein.

25

105. The method according to claim 104 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

106. Isolated nucleic acid encoding a *CYP52A5A* protein having the amino acid sequence set forth in SEQ ID NO: 100.

30

107. Isolated nucleic acid comprising a coding region defined by nucleotides 1103-2656 as set forth in SEQ ID NO: 90.

108. Isolated nucleic acid according to claim 107 comprising the nucleotide sequence as set forth in SEQ ID NO: 90.

5 109. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 100.

110. A vector comprising a nucleotide sequence encoding *CYP52A5A* protein including an amino acid sequence as set forth in SEQ ID NO: 100.

10 111. A vector according to claim 110 wherein the nucleotide sequence is set forth in nucleotides 1103-2656 OF SEQ ID NO: 90.

112. A vector according to claim 110 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.
15

113. A host cell transfected or transformed with the nucleic acid of claim 106.

114. A host cell according to claim 113 wherein the host cell is a yeast cell.

20 115. A host cell according to claim 114 wherein the yeast cell is a *Candida sp.*

116. A host cell according to claim 115 wherein the *Candida sp.* is *Candida tropicalis*.

25 117. A host cell according to claim 116 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

118. A host cell according to claim 117 wherein the *Candida tropicalis* is H5343 ura-.
30

119. A method of producing a *CYP52A5A* protein including an amino acid sequence as set forth in SEQ ID NO: 100 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 100; and

b) culturing the cell under conditions favoring the expression of the protein.

5 120. The method according to claim 119 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

 121. Isolated nucleic acid encoding a *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101.

10

 122. Isolated nucleic acid comprising a coding region defined by nucleotides 1142-2695 as set forth in SEQ ID NO: 91.

 123. Isolated nucleic acid according to claim 122 comprising the nucleotide
15 sequence as set forth in SEQ ID NO: 91.

 124. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 101.

20 125. A vector comprising a nucleotide sequence encoding *CYP52A5B* protein including the amino acid sequence as set forth in SEQ ID NO: 101.

 126. A vector according to claim 125 wherein the nucleotide sequence is set forth in nucleotides 1142-2695 of SEQ ID NO: 91.

25

 127. A vector according to claim 125 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

 128. A host cell transfected or transformed with the nucleic acid of claim 121.

30

 129. A host cell according to claim 128 wherein the host cell is a yeast cell.

130. A host cell according to claim 129 wherein the yeast cell is a *Candida sp.*

131. A host cell according to claim 130 wherein the *Candida sp.* is *Candida tropicalis*.

5

132. A host cell according to claim 131 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

133. A host cell according to claim 132 wherein the *Candida tropicalis* is H5343
10 ura-.

134. A method of producing a *CYP52A5B* protein including an amino acid sequence as set forth in SEQ ID NO: 101 comprising:

- 15 a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 101; and
b) culturing the cell under conditions favoring the expression of the protein.

135. The method according to claim 134 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.
20

136. Isolated nucleic acid encoding a *CYP52A8A* protein having the amino acid sequence set forth in SEQ ID NO: 102.

137. Isolated nucleic acid comprising a coding region defined by nucleotides 464-
25 2002 as set forth in SEQ ID NO: 92.

138. Isolated nucleic acid according to claim 137 comprising the nucleotide sequence as set forth in SEQ ID NO: 92.

139. Isolated protein comprising an amino acid sequence as set forth in SEQ ID
30 NO: 102.

140. A vector comprising a nucleotide sequence encoding *CYP52A8A* protein including an amino acid sequence as set forth in SEQ ID NO: 102.

141. A vector according to claim 140 wherein the nucleotide sequence is set forth
5 in nucleotides 464-2002 of SEQ ID NO: 92.

142. A vector according to claim 140 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

10 143. A host cell transfected or transformed with the nucleic acid of claim 136.

144. A host cell according to claim 143 wherein the host cell is a yeast cell.

145. A host cell according to claim 144 wherein the yeast cell is a *Candida sp.*

15 146. A host cell according to claim 145 wherein the *Candida sp.* is *Candida tropicalis*.

147. A host cell according to claim 146 wherein the *Candida tropicalis* is
20 *Candida tropicalis* 20336.

148. A host cell according to claim 147 wherein the *Candida tropicalis* is H5343
ura-.

25 149. A method of producing a *CYP52A8A* protein including an amino acid sequence as set forth in SEQ ID NO: 102 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 102; and

b) culturing the cell under conditions favoring the expression of the protein.

30 150. The method according to claim 149 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

151. Isolated nucleic acid encoding a *CYP52A8B* protein having the amino acid sequence set forth in SEQ ID NO: 103.

152. Isolated nucleic acid comprising a coding region defined by nucleotides
5 1017-2555 as set forth in SEQ ID NO: 93.

153. Isolated nucleic acid according to claim 152 comprising the nucleotide sequence as set forth in SEQ ID NO: 93.

10 154. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 103.

155. A vector comprising a nucleotide sequence encoding *CYP52A8B* protein including an amino acid sequence as set forth in SEQ ID NO: 103.

15 156. A vector according to claim 155 wherein the nucleotide sequence is set forth in nucleotides 1017-2555 of SEQ ID NO: 93.

157. A vector according to claim 155 wherein the vector is selected from the
20 group consisting of plasmid, phagemid, phage and cosmid.

158. A host cell transfected or transformed with the nucleic acid of claim 151.

159. A host cell according to claim 158 wherein the host cell is a yeast cell.
25

160. A host cell according to claim 159 wherein the yeast cell is a *Candida sp.*

161. A host cell according to claim 160 wherein the *Candida sp.* is *Candida tropicalis*.
30

162. A host cell according to claim 161 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

163. A host cell according to claim 162 wherein the *Candida tropicalis* is H5343 ura-.

164. A method of producing a *CYP52A8B* protein including an amino acid
5 sequence as set forth in SEQ ID NO: 103 comprising:
a) transforming a suitable host cell with a DNA sequence that encodes the protein
having the amino acid sequence as set forth in SEQ ID NO: 103; and
b) culturing the cell under conditions favoring the expression of the protein.

10 165. The method according to claim 164 wherein the step of culturing the cell
comprises adding an organic substrate to media containing the cell.

166. Isolated nucleic acid encoding a *CYP52D4A* protein having the amino acid
sequence set forth in SEQ ID NO: 104.

15 167. Isolated nucleic acid comprising a coding region defined by nucleotides 767-
2266 as set forth in SEQ ID NO: 94.

168. Isolated nucleic acid according to claim 167 comprising the nucleotide
20 sequence as set forth in SEQ ID NO: 94.

169. Isolated protein comprising an amino acid sequence as set forth in SEQ ID
NO: 104.

25 170. A vector comprising a nucleotide sequence encoding *CYP52D4A* protein
including an amino acid sequence as set forth in SEQ ID NO: 104.

171. A vector according to claim 170 wherein the nucleotide sequence is set forth
in nucleotides 767-2266 of SEQ ID NO: 94.

30 172. A vector according to claim 170 wherein the vector is selected from the
group consisting of plasmid, phagemid, phage and cosmid.

173. A host cell transfected or transformed with the nucleic acid of claim 166.

174. A host cell according to claim 173 wherein the host cell is a yeast cell.

5 175. A host cell according to claim 174 wherein the yeast cell is a *Candida sp.*

176. A host cell according to claim 175 wherein the *Candida sp.* is *Candida tropicalis*.

10 177. A host cell according to claim 176 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

178. A host cell according to claim 177 wherein the *Candida tropicalis* is H5343 ura-.

15

179. A method of producing a *CYP52D4A* protein including an amino acid sequence as set forth in SEQ ID NO: 104 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 104; and

20 b) culturing the cell under conditions favoring the expression of the protein.

180. The method according to claim 179 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

25 181. A method for discriminating members of a gene family by quantifying the amount of target mRNA in a sample comprising:

a) providing an organism containing a target gene;

b) culturing the organism with an organic substrate which causes upregulation in the activity of the target gene;

30 c) obtaining a sample of total RNA from the organism at a first point in time;

- d) combining at least a portion of the sample of the total RNA with a known amount of competitor RNA to form an RNA mixture, wherein the competitor RNA is substantially similar to the target mRNA but has a lesser number of nucleotides compared to the target mRNA;
- 5 e) adding reverse transcriptase to the RNA mixture in a quantity sufficient to form corresponding target DNA and competitor DNA;
- f) conducting a polymerase chain reaction in the presence of at least one primer specific for at least one substantially non-homologous region of the target DNA within the gene family, the primer also specific for the competitor DNA;
- 10 g) repeating steps (c-f) using increasing amounts of the competitor RNA while maintaining a substantially constant amount of target RNA;
- (h) determining the point at which the amount of target DNA is substantially equal to the amount of competitor DNA;
- (i) quantifying the results by comparing the ratio of the concentration of
- 15 unknown target to the known concentration of competitor; and
- (j) obtaining a sample of total RNA from the organism at another point in time and repeating steps (d-i).

182. A method according to claim 181 wherein the target gene is selected from

20 the group consisting of a *CPR* gene and a *CYP* gene.

183. A method according to claim 182 wherein the *CPR* gene is selected from the group consisting of a *CPRA* gene (SEQ ID NO: 81) and a *CPRB* gene (SEQ ID NO: 82).

25 184. A method according to claim 182 wherein the *CYP* gene is selected from the group consisting of *CYP52A1A* gene (SEQ ID NO: 85), *CYP52A2A* gene (SEQ ID NO: 86), *CYP52A2B* gene (SEQ ID NO: 87), *CYP52A3A* gene (SEQ ID NO: 88), *CYP52A3B* gene (SEQ ID NO: 89), *CYP52A5A* gene (SEQ ID NO: 90), *CYP52A5B* gene (SEQ ID NO: 91), *CYP52A8A* gene (SEQ ID NO: 92), *CYP52A8B* gene (SEQ ID NO: 93) and *CYP52D4A* gene (SEQ ID NO:

30 94).

185. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CPRA* genes;

b) increasing, in the host cell, the number of *CPRA* genes which encode a *CPRA* protein having the amino acid sequence as set forth in SEQ ID NO: 83;

5 c) culturing the host cell in media containing an organic substrate which upregulates the *CPRA* gene, to effect increased production of dicarboxylic acid.

186. A method for increasing the production of a *CPRA* protein having an amino acid sequence as set forth in SEQ ID NO: 83 comprising:

10 a) transforming a host cell having a naturally occurring amount of *CPRA* protein with an increased copy number of a *CPRA* gene that encodes the *CPRA* protein having the amino acid sequence as set forth in SEQ ID NO: 83; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CPRA* gene.

15

187. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CPRB* genes;

b) increasing, in the host cell, the number of *CPRB* genes which encode a *CPRB* protein having the amino acid sequence as set forth in SEQ ID NO: 84;

20 c) culturing the host cell in media containing an organic substrate which upregulates the *CPRB* gene, to effect increased production of dicarboxylic acid.

188. A method for increasing the production of a *CPRB* protein having an amino acid sequence as set forth in SEQ ID NO: 84 comprising:

25 a) transforming a host cell having a naturally occurring amount of *CPRB* protein with an increased copy number of a *CPRB* gene that encodes the *CPRB* protein having the amino acid sequence as set forth in SEQ ID NO: 84; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CPRB* gene.

30

189. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A1A* genes;

b) increasing, in the host cell, the number of *CYP52A1A* genes which encode a *CYP52A1A* protein having the amino acid sequence as set forth in SEQ ID NO: 95;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2A* gene, to effect increased production of dicarboxylic acid.

5

190. A method for increasing the production of a *CYP52A1A* protein having an amino acid sequence as set forth in SEQ ID NO: 95 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A1A* protein with an increased copy number of a *CYP52A1A* gene that encodes the *CYP52A1A* protein
10 having the amino acid sequence as set forth in SEQ ID NO: 95; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A1A* gene.

191. A method for increasing production of a dicarboxylic acid comprising:

15 a) providing a host cell having a naturally occurring number of *CYP52A2A* genes;

b) increasing, in the host cell, the number of *CYP52A2A* genes which encode a *CYP52A2A* protein having the amino acid sequence as set forth in SEQ ID NO: 96;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2A* gene, to effect increased production of dicarboxylic acid.

20

192. A method for increasing the production of a *CYP52A2A* protein having an amino acid sequence as set forth in SEQ ID NO: 96 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A2A* protein with an increased copy number of a *CYP52A2A* gene that encodes the *CYP52A2A* protein
25 having the amino acid sequence as set forth in SEQ ID NO: 96; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A2A* gene.

193. A method for increasing production of a dicarboxylic acid comprising:

30 a) providing a host cell having a naturally occurring number of *CYP52A2B* genes;

b) increasing, in the host cell, the number of *CYP52A2B* genes which encode a *CYP52A2B* protein having the amino acid sequence as set forth in SEQ ID NO: 97;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2B* gene, to effect increased production of dicarboxylic acid.

194. A method for increasing the production of a *CYP52A2B* protein having an amino acid sequence as set forth in SEQ ID NO: 97 comprising:

- a) transforming a host cell having a naturally occurring amount of *CYP52A2B* protein with an increased copy number of a *CYP52A2B* gene that encodes the *CYP52A2B* protein having the amino acid sequence as set forth in SEQ ID NO: 97; and
- b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A2B* gene.

195. A method for increasing production of a dicarboxylic acid comprising:

- a) providing a host cell having a naturally occurring number of *CYP52A3A* genes;
- b) increasing, in the host cell, the number of *CYP52A3A* genes which encode a *CYP52A3A* protein having the amino acid sequence as set forth in SEQ ID NO: 98;
- c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A3A* gene, to effect increased production of dicarboxylic acid.

196. A method for increasing the production of a *CYP52A3A* protein having an amino acid sequence as set forth in SEQ ID NO: 98 comprising:

- a) transforming a host cell having a naturally occurring amount of *CYP52A3A* protein with an increased copy number of a *CYP52A3A* gene that encodes the *CYP52A3A* protein having the amino acid sequence as set forth in SEQ ID NO: 98; and
- b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A3A* gene.

197. A method for increasing production of a dicarboxylic acid comprising:

- a) providing a host cell having a naturally occurring number of *CYP52A3B* genes;
- b) increasing, in the host cell, the number of *CYP52A3B* genes which encode a *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99;
- c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A3B* gene, to effect increased production of dicarboxylic acid.

198. A method for increasing the production of a *CYP52A3B* protein having an amino acid sequence as set forth in SEQ ID NO: 99 comprising:

- 5 a) transforming a host cell having a naturally occurring amount of *CYP52A3B* protein with an increased copy number of a *CYP52A3B* gene that encodes the *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99; and
- b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A3B* gene.

199. A method for increasing production of a dicarboxylic acid comprising:

- 10 a) providing a host cell having a naturally occurring number of *CYP52A5A* genes;
- b) increasing, in the host cell, the number of *CYP52A5A* genes which encode a *CYP52A5A* protein having the amino acid sequence as set forth in SEQ ID NO: 100;
- c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A5A* gene, to effect increased production of dicarboxylic acid.

15

200. A method for increasing the production of a *CYP52A5A* protein having an amino acid sequence as set forth in SEQ ID NO: 100 comprising:

- a) transforming a host cell having a naturally occurring amount of *CYP52A5A* protein with an increased copy number of a *CYP52A5A* gene that encodes the *CYP52A5A* protein
- 20 having the amino acid sequence as set forth in SEQ ID NO: 100; and
- b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A5A* gene.

201. A method for increasing production of a dicarboxylic acid comprising:

- 25 a) providing a host cell having a naturally occurring number of *CYP52A5B* genes;
- b) increasing, in the host cell, the number of *CYP52A5B* genes which encode a *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101;
- c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A5B* gene, to effect increased production of dicarboxylic acid.

30

202. A method for increasing the production of a *CYP52A5B* protein having an amino acid sequence as set forth in SEQ ID NO: 101 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A5B* protein with an increased copy number of a *CYP52A5B* gene that encodes the *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101; and

b) culturing the cell and thereby increasing expression of the protein compared
5 with that of a host cell containing a naturally occurring copy number of the *CYP52A5B* gene.

203. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A8A* genes;

b) increasing, in the host cell, the number of *CYP52A8A* genes which encode a
10 *CYP52A8A* protein having the amino acid sequence as set forth in SEQ ID NO: 102;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A8A* gene, to effect increased production of dicarboxylic acid.

204. A method for increasing the production of a *CYP52A8A* protein having an amino acid sequence as set forth in SEQ ID NO: 102 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A8A* protein with an increased copy number of a *CYP52A8A* gene that encodes the *CYP52A8A* protein having the amino acid sequence as set forth in SEQ ID NO: 102; and

b) culturing the cell and thereby increasing expression of the protein compared
20 with that of a host cell containing a naturally occurring copy number of the *CYP52A8A* gene.

205. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A8B* genes;

b) increasing, in the host cell, the number of *CYP52A8B* genes which encode a
25 *CYP52A8B* protein having the amino acid sequence as set forth in SEQ ID NO: 103;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A8B* gene, to effect increased production of dicarboxylic acid.

206. A method for increasing the production of a *CYP52A8B* protein having an amino acid sequence as set forth in SEQ ID NO: 103 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A8B* protein with an increased copy number of a *CYP52A8B* gene that encodes the *CYP52A8B* protein having the amino acid sequence as set forth in SEQ ID NO: 103; and

b) culturing the cell and thereby increasing expression of the protein compared
5 with that of a host cell containing a naturally occurring copy number of the *CYP52A8B* gene.

207. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52D4A* genes;

b) increasing, in the host cell, the number of *CYP52D4A* genes which encode a
10 *CYP52D4A* protein having the amino acid sequence as set forth in SEQ ID NO: 104;

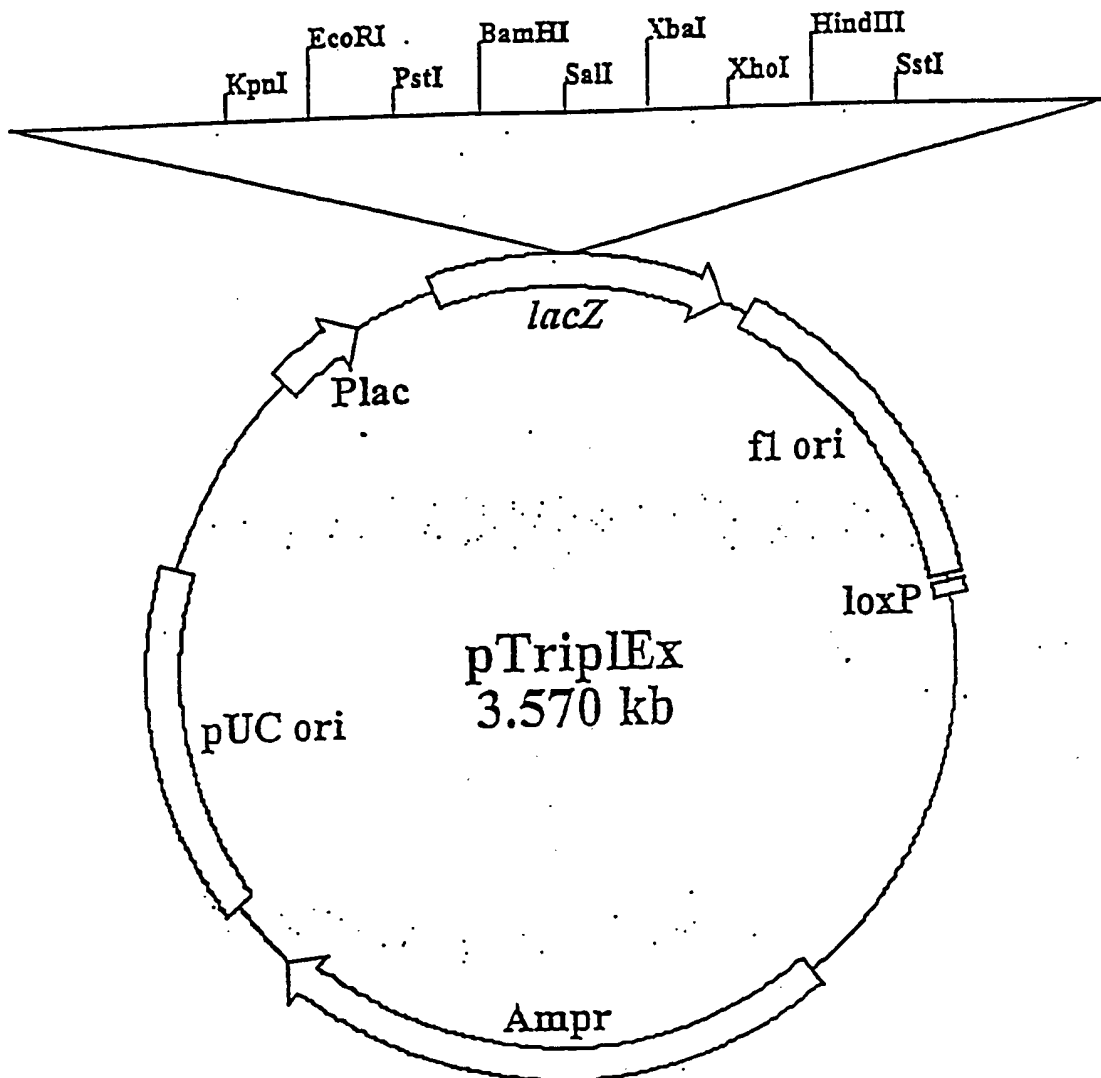
c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52D4A* gene, to effect increased production of dicarboxylic acid.

208. A method for increasing the production of a *CYP52D4A* protein having an
15 amino acid sequence as set forth in SEQ ID NO: 104 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52D4A* protein with an increased copy number of a *CYP52D4A* gene that encodes the *CYP52D4A* protein having the amino acid sequence as set forth in SEQ ID NO: 104; and

b) culturing the cell and thereby increasing expression of the protein compared
20 with that of a host cell containing a naturally occurring copy number of the *CYP52D4A* gene.

209. A method for discriminating members of a gene family according to claim
181 wherein culturing the organism with the organic substrate is accomplished in a fermentor.

*Figure 1*

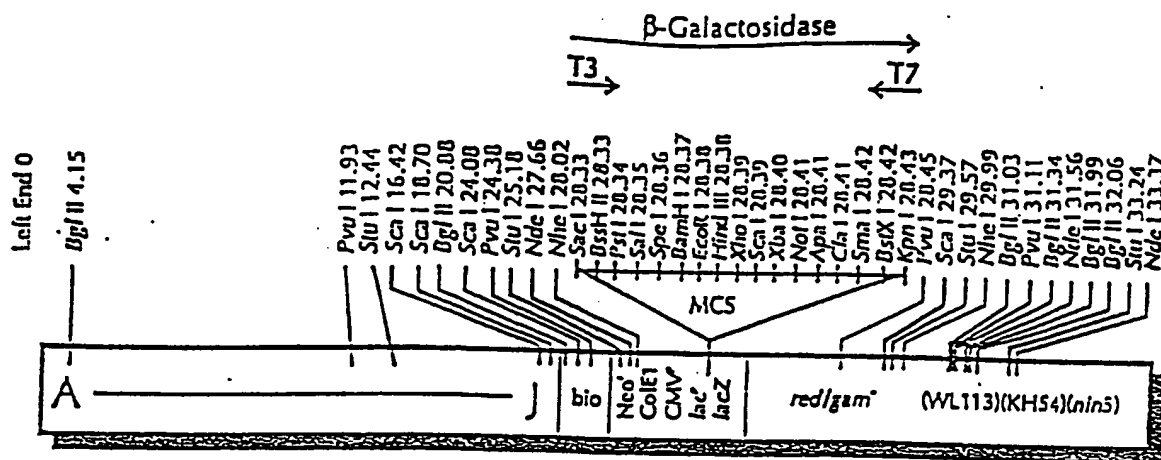


Figure 2A

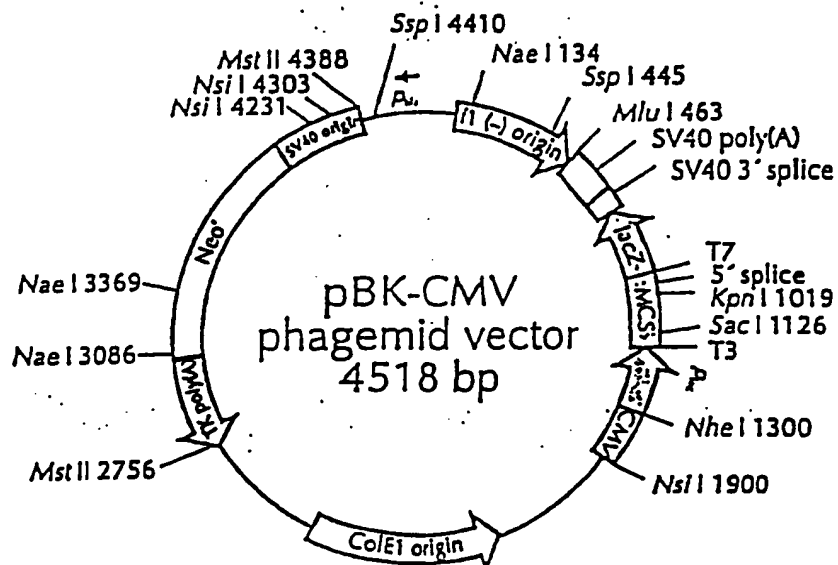


Figure 2B

QC-RT-PCR primers for the 5' coding sequence of
Candida tropicalis 20336 P450CYP52A5A

```

5' ATGATTGACAACTCCTAGAATATGGTAT GTGTTGTGCCAGTGTGTACATCATCAA CAACTCCTTGCCATACACAAGACTCGCGTC 3' 90
3' TACTAACTTGTGTGAGGATCTTATAACCAT CAGCAACACGGTCACAAATGTAGTAGTTT GTTGAGGAAAGTATGTGTTTCTGAGCGCAG 5'

5' TTGATGAAAAAGTTGGGTGCTGCTCCAGTC ACNACAGTTGTACGACACGCTTTCGGT ATGTCATGGGATGGAAGCTCTCCAGTTC 3' 180
3' AACTACTTITTTCAACCACGACGAGGTCAG TGTITGTTCAACATGCTGTTCGGAAGGCA TAGCACTACCTACCTTCGAGAGGTCAG 5'

Forward Primer 7551-577
5' AAGTAAAGTGGCTGGCTCAAGATACAC GATTACAGTTTGCCACTCCAGAACCA AGGTGGGCACTACGTCACTATTCCTTTC 3' 270
3' TTCTTTCCTCCGACCGGTTCCTCATGTG CTATGTTCAACTGGTGGGTCTTGGT TCACACCGTGGATGCAGTCATAGGAAG 5'

5' GGCACCAGGATCGTCGTACCAAGATCCA GAGATATCAAAGCTATTTGGCAACCCAG TTGGTGATTTTCTTTGGCAAGAGGCAC 3' 360
3' CCGTGGTCTAGGACACTGGTTTCTAGGT CTCTATAGTTTGTATAAACCGTTGGGC AAACCACTAAAGAAACCGTCTCCGTG 5'

5' ACTCTTTTAAAGCTTTGTTAGGTGATGG ATCTTCATTTGGACGGGGAAGCTGGAG CACAGCAGGCCCTGTGTGACCCAGTTT 3' 450
3' TGGAAAATTCCGAAACATCCTACCC TAGAAGTGAACCTCCGCTCCGACCTTC GTGTCGTCTCGGTACACTCTGGGTCAA 5'

Reverse Primer 7541-574
5' GCGAGAGACAAAGTGCCTCATGTGACGTG TTGGAACACACTTCAGTTGTGAAGAG CATATTCTTAAGCACAAGGTGATACTTT 3' 540
3' CGGTCTCTTGTTCACGAGTACATGCAAG AACCTTGGTGTGAAGGTCAACACTTCTC GTATAGAAATTCGTGTTCCACTATGAA 5'

```

Figure 3

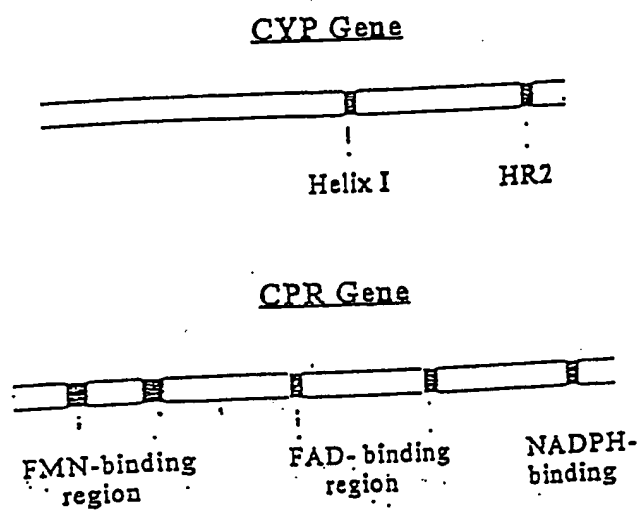


Figure 4

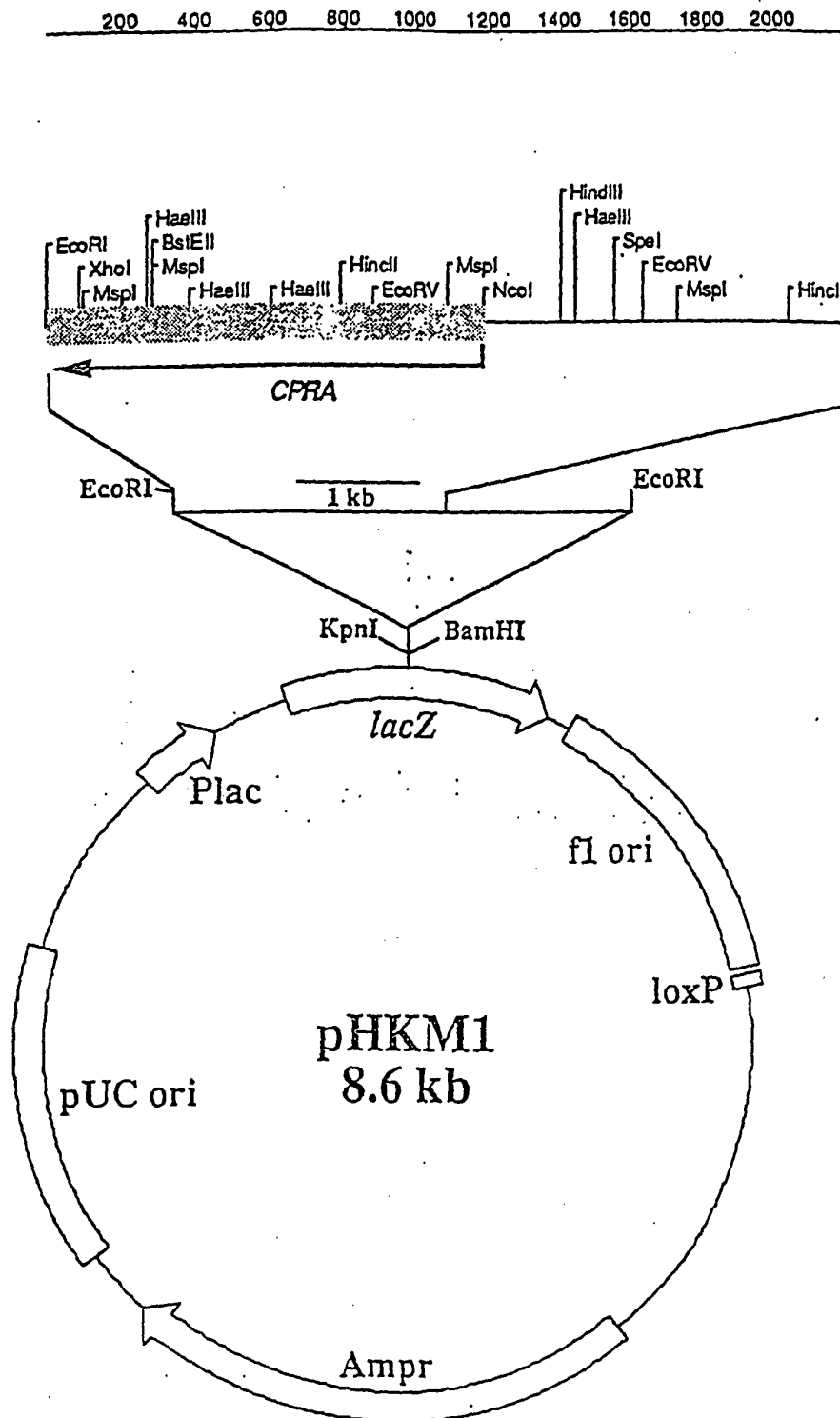


Figure 5

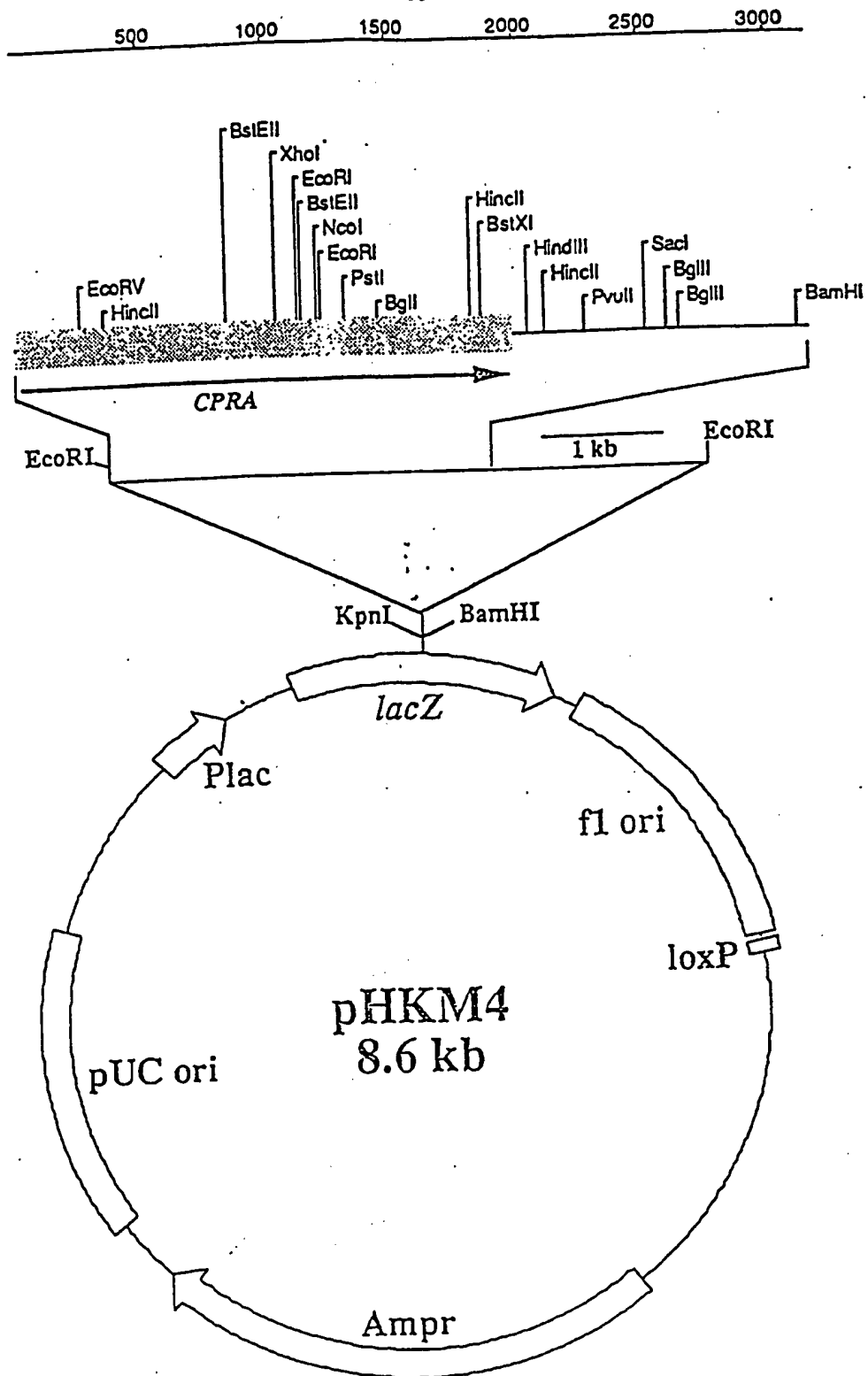


Figure 6

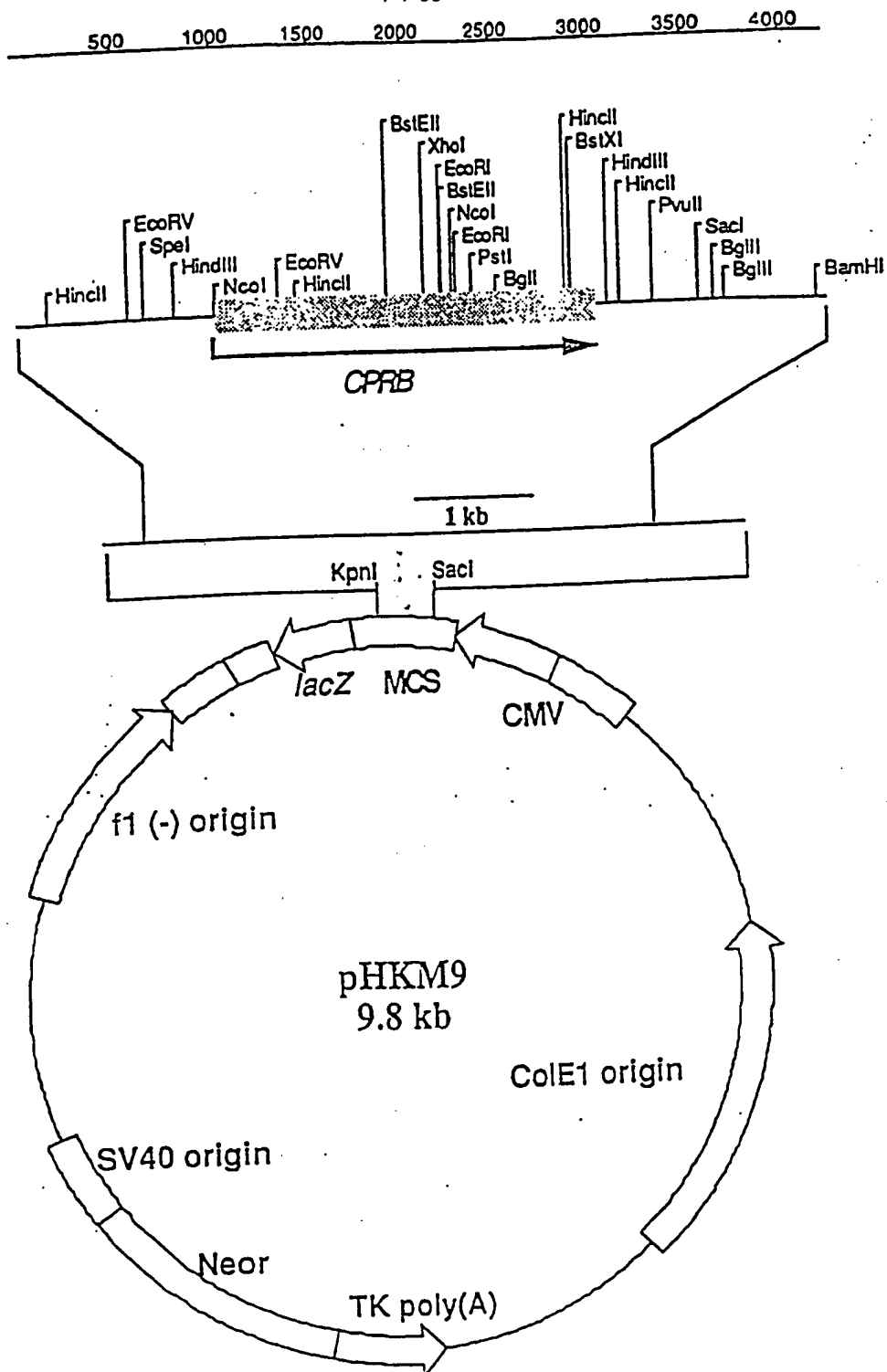


Figure 7

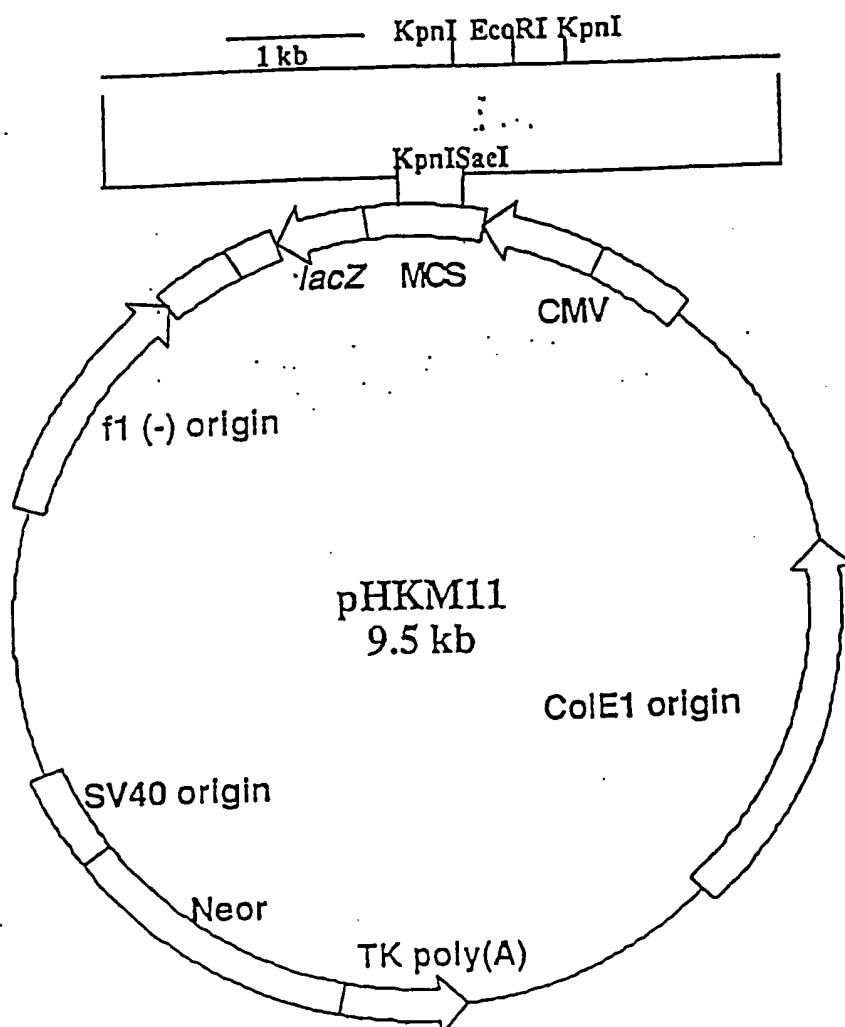
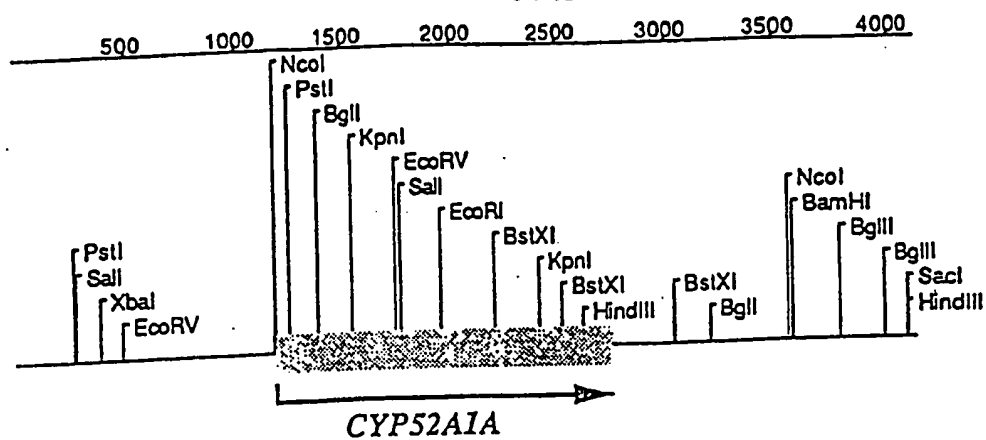


Figure 8

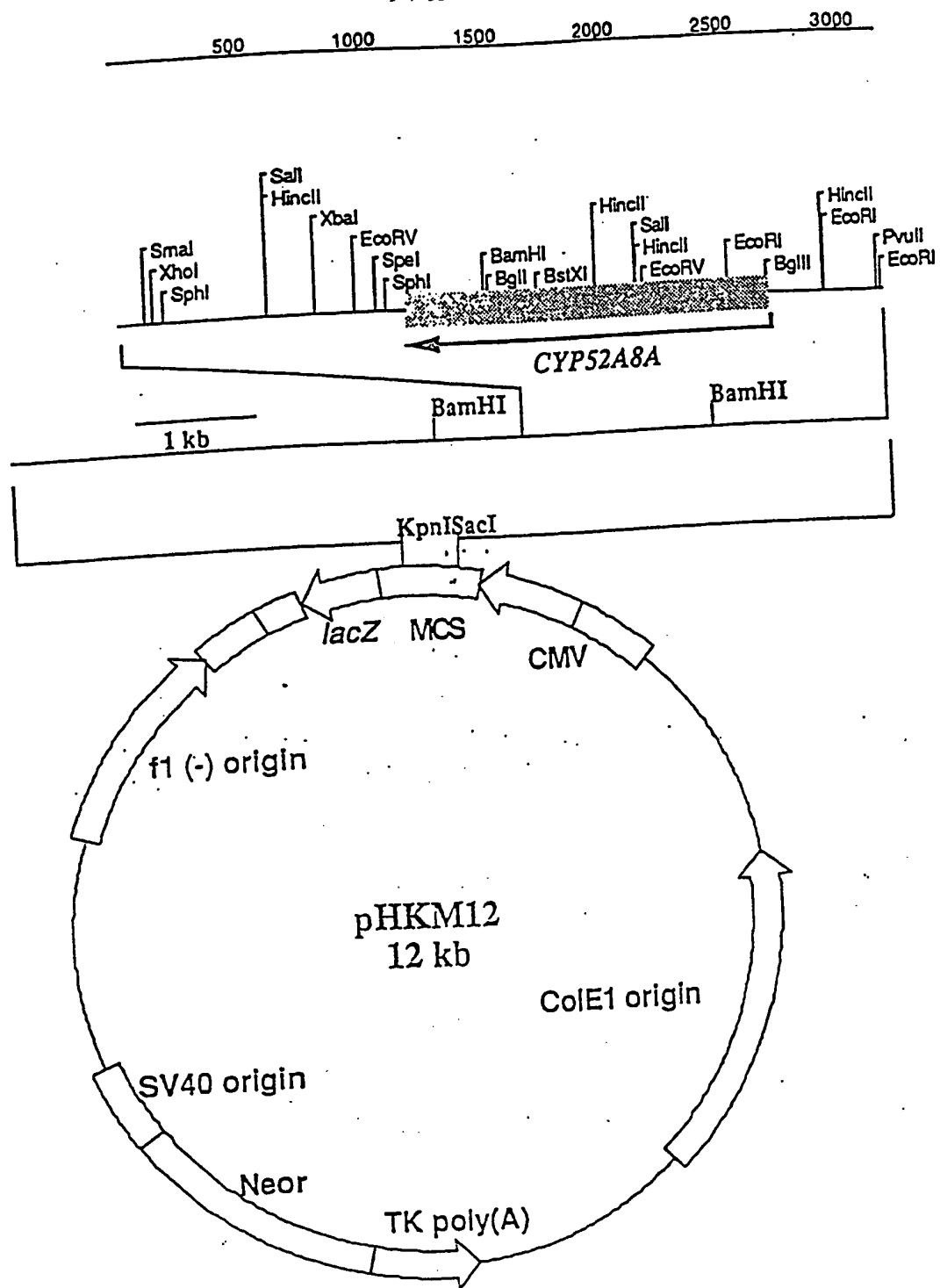


Figure 9

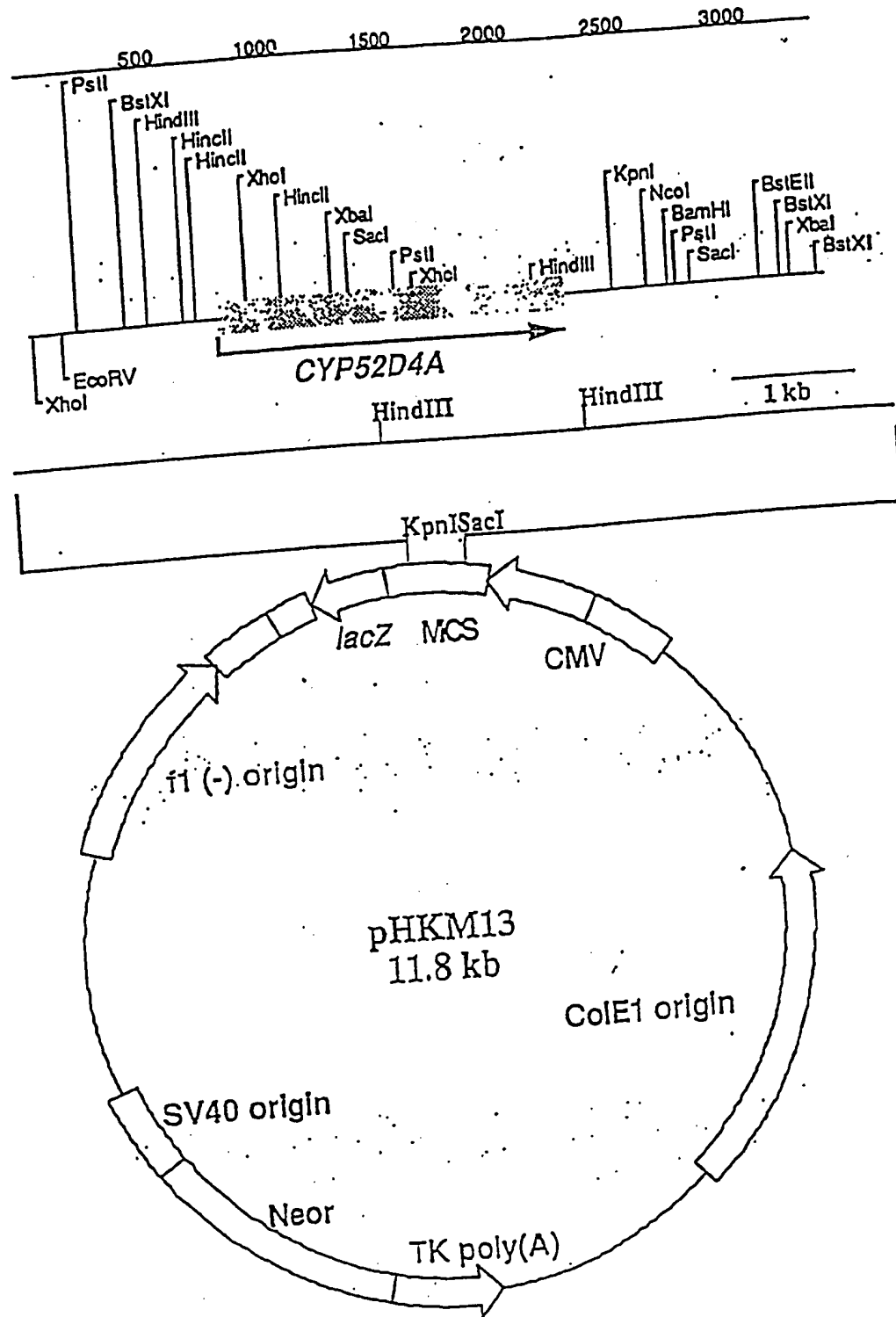


Figure 10

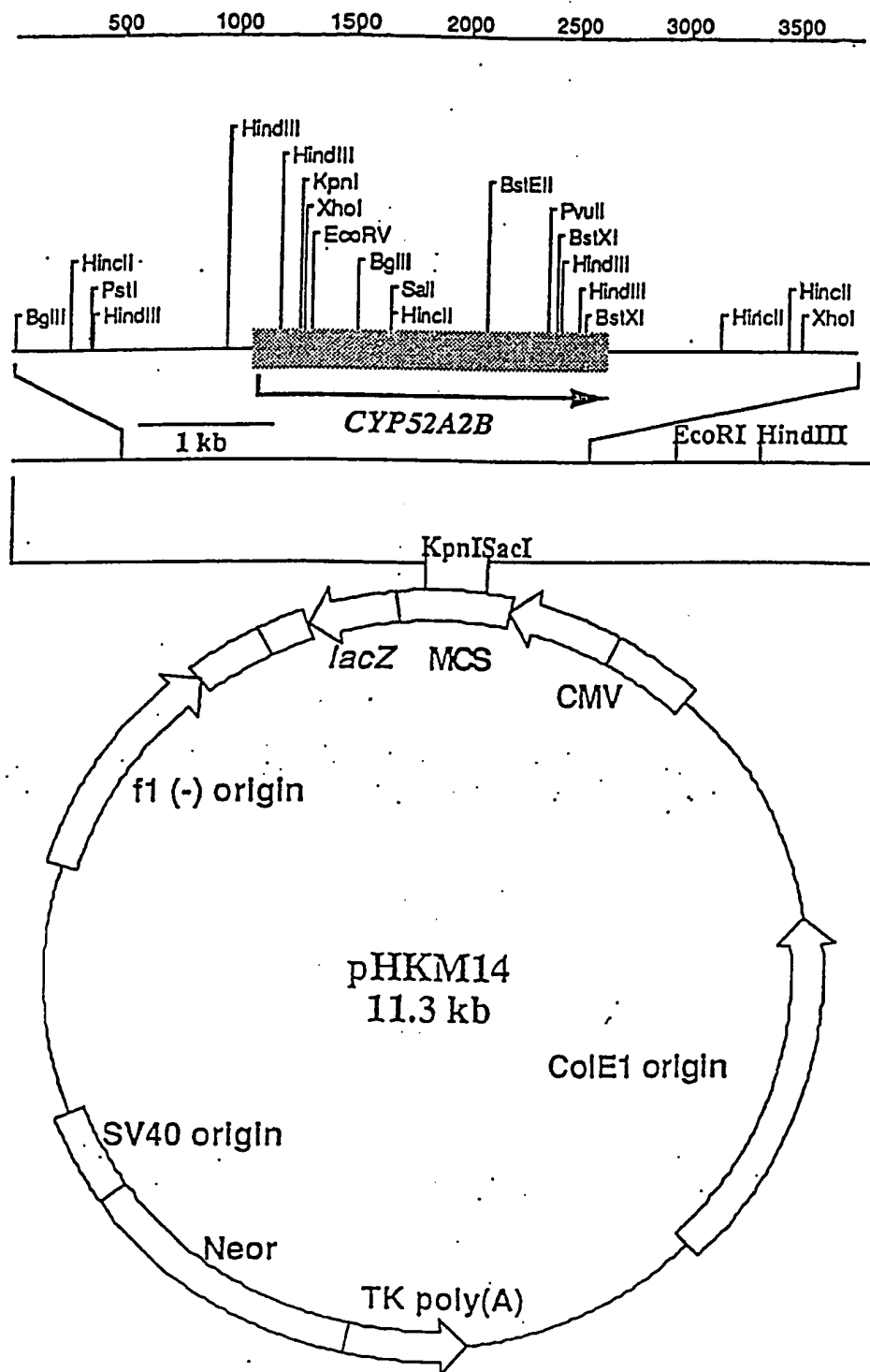


Figure 11

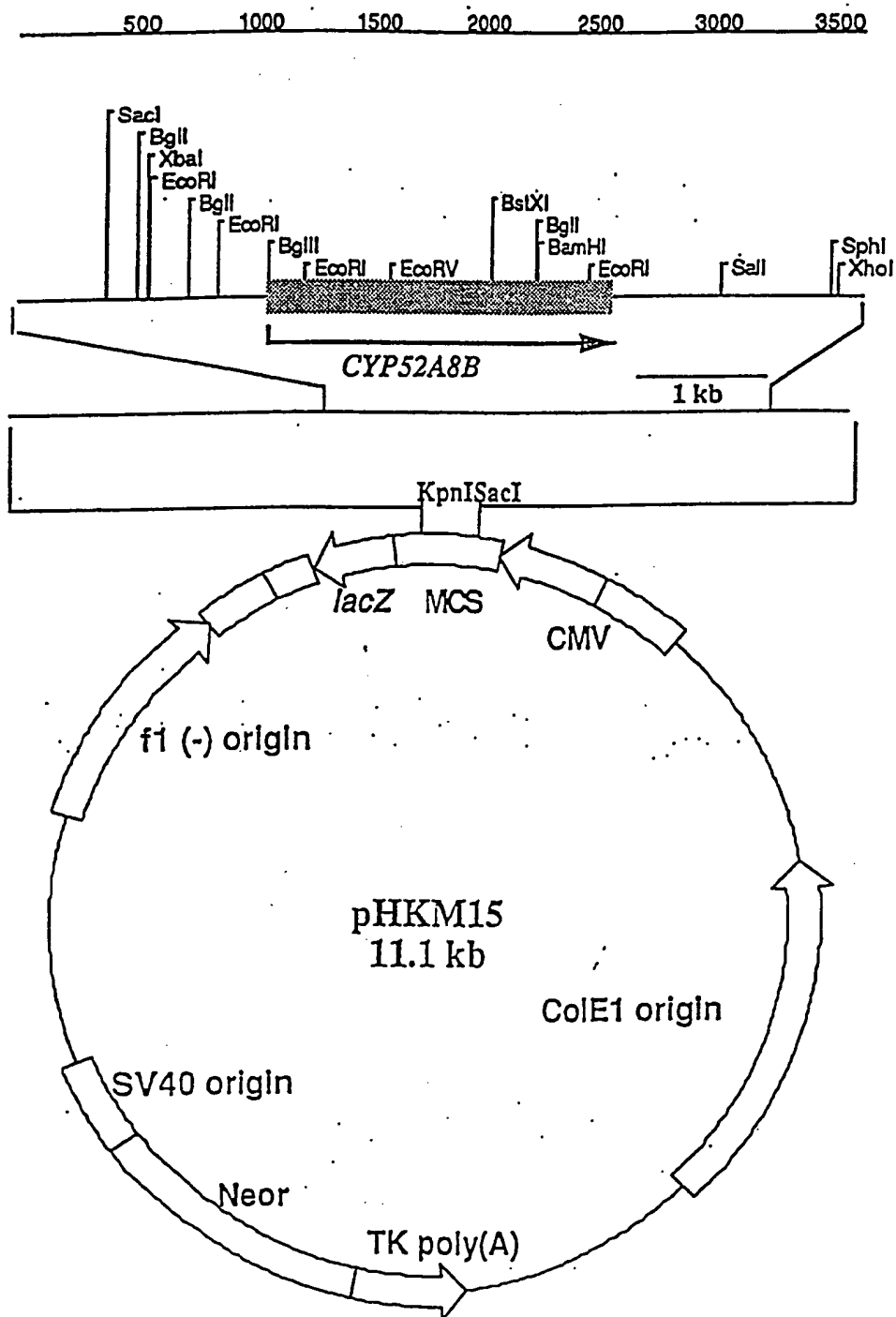


Figure 12

C. tropicalis 20336 CPR Allele DNA Alignment of DS Sequence

CPRA	1		CATCA	5
CPRB	1	TATATGATATATGATATATCTTCTGTGTAATTATTATTCGTATTCGTAATACTTACTACATTTTTTTT		70
CPRA	6	AGATCATCTATGGGGATAATTA-----CGACAGCAACATTGCAGAAAGAGCGTTGGTCACAATCGAAAGA		70
CPRB	71	TCITTTATTTATGAAGAAAAGGAGAGTTCGTAGATTGAGTTGAGTAGAATAGGCTGTTTGTGCATACGGGGA		140
CPRA	71	GCCTATG-GCGTTGCCGTGCTGTGAGGCAATGACAGCAC--CAACAATAACGATGGTCCAGTGAAGAGC		137
CPRB	141	GCAGAGGAGAGTATCCGACGAGGAGGAACGGGTGAAATTCATCTATGCTGTTGCGTCTGTACTGTAC		210
CPRA	138	CTTCAGAACAGTCCATTGTTGACGCT--TAAGGCAAGGATAATTACGTGGGGCAAGGAACGGGAATTA		205
CPRB	211	TGTAAATCTTAGATTTCCTAGAGGTTGTTCTAGCAATAAAGTGTTCAAGATACAATTTTACAGGCAAG		280
CPRA	206	GTTATGGGGGATCAAA--AGCGGAAGATTGTGTGTGCTTGTGGGTTTTTCTTTATTTTTCATATGAT		273
CPRB	281	GGTAAAGGATCAACTGATTAGCGGAAGATTGTTGTTGCCGTGGGGTTCTT---TTATTTTTCATATGAT		347
CPRA	274	TTCTTTGCGCAAGTAACATGTGCCAATTAGTTTGTGATTAGCGTGCC-CCACAATTGGCATCGTGGAGC		342
CPRB	348	TTCTTTGCGCAGTAACATGTGCCAATCTAGTTTATGATTAGCGTACCTCCACAATTGGCATCTTGGAGC		417
CPRA	343	GGCGTGTTTTGTATACCCCAAGTCTTAACTAGCTCCACAGTCTCGACGGTGTCTCGACGATGCTCTTCTT		412
CPRB	418	GGCGTGTTTTGTCTTACCCCAAGCCTTATTAGTTCCACAGTCTCGACGGTGTCTCGCGCATGCTCTCTC		487
CPRA	413	CCACCCCTCCCATGAATCATTCAAAGTTGTTGGGGGATCTCCACCAAGGGCACCGAGTTAATGCTTATG		482
CPRB	488	CCACCCCTCGCAGGAATCATTGAAAGTTGTTGGGGGATCTCCTCC-----CGAGTTTATGTTATG		548
CPRA	483	TTTCTCCCACTTTGTTGTGATTGGGGTAGTCTAGTGAGTTGGAGATTTCTTTTTCGCAAGGTGCTC		552
CPRB	549	CTTTTCCCACTTTGTTGTGATTGGGGTAGCGTAGTGAGTTTGTGATTTCTTTTTCGCAAGGTGCTC		617
CPRA	553	CGATATCGAAATTTGATGAATATAGAGAGAAAGCCAGATCAGCACAGTAGATTGCTTTGTAGTTAGAGAT		622
CPRB	618	CGATATCGAAGTTTGTATGAATATAG----GAGCCAGATCAGCATGTTATATTGCTTTGTAGATAGAGAT		683
CPRA	623	GTTGAACAGCAACTAGTTGAATTACAGCCCACTTGACAGCAAGTGCGAGTGAGCTGTAAACGATGCAG		692
CPRB	664	GTTGAACACCACTAGCTGAATTACACCAACCGCT-----AAACGATGCAG		730
CPRA	693	CCAGAGTGTCAACCACTGACGTTGGGTGGAGTTGTTGTTGTTGTTGTTGGCAGGGCCATATTGCTAA		762
CPRB	731	ACAGGGTGTCAACGCCAATGACGTTGGGTGGAGTTG-----TTGTTGGCAGGGCCATATTGCTAA		791
CPRA	763	ACGAAGACAGTAGCACAAAACCCAGGCTTAAGAAACAAAATAAAAAAATTCATACGACAATCCAAAG		832
CPRB	792	ACGAAGAGAGTAGCACAAAACCCAGGCTTAAGAAACAAAATAAAAAAATTCATACGACAATCCAAAG		858
CPRA	833	CCATTGATTACATAAT--CAACAG-TAAGACAGAAAAAATTTCACATTTCAGAGTTCCCTTTTCTCT		899
CPRB	859	CCATTACATAATCAACAGCGACAAATGAGACAGAAAAAATTTCACATTTCAGAGTTCCCTTTTCTCT		928
CPRA	900	ATTACTCTTTTTTTTCTCTTTCTTTCTTT-----CTTCTCTCTGTTTTCTTACTTTATCAGTCTTTTA		962
CPRB	929	ATTACTCTTTTTTTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT		998
CPRA	953	CTTGTTTTTGCAATTCCTCATCT		1032
CPRB	999	CTTGTTTTTGCAATTCCTCATCT		1059

Figure 13A

CPRA 1033 GTCATCATAACATTGGTGGTGGCTGTAGCCGCCCTATTTTGTCTAAGAACCAGTTCCCTTGATCAGCCCCAGG 1102
 CPRB 1060 GTCATCATAACATTGGTGGTGGCTGTAGCCGCCCTATTTTGTCTAAGAACCAGTTCCCTTGATCAGCCCCAGG 1129

 CPRA 1103 ACACCGGGTTCTCTCAACACGGACAGCGGAAGCAACTCCAGAGACGCTCTTGCTGACATTGAAGAAGAAATA 1172
 CPRB 1130 ACACCGGGTTCTCTCAACACGGACAGCGGAAGCAACTCCAGAGACGCTCTTGCTGACATTGAAGAAGAAATA 1199

 CPRA 1173 TAAAAACACGTTGTGTGTGTGTGGGTCCAGACGGGTACGGCAGAAAGATTACGCCAACAAATTGTCCAGA 1242
 CPRB 1200 TAAAAACACGTTGTGTGTGTGTGGGTCCAGACGGGTACGGCAGAAAGATTACGCCAACAAATTGTCCAGA 1269

 CPRA 1243 GAATTGCACTCCAGATTGTGGCTTGAJAAACGATGGTTGCAGATTTCGCTGATTACGATTGGGATAAATTTCG 1312
 CPRB 1270 GAATTGCACTCCAGATTGTGGCTTGAJAAACGATGGTTGCAGATTTCGCTGATTACGATTGGGATAAATTTCG 1339

 CPRA 1313 GAGATATCACCAGAGACATCTTGGTGTGTTCATTGTTGCCACCTATGGTGAGGGTGAACCTACCGATA 1382
 CPRB 1340 GAGATATCACCAGAGATATCTTGGTGTGTTCATTGTTGCCACCTACGGTGAGGGTGAACCTACCGATA 1409

 CPRA 1383 TGCCGACGAGTTCCACAC CTGGTTGACTGAAGAAGCTGACACTTTGAGTACCTTGAATACACCGTGTTC 1452
 CPRB 1410 TGCCGACGAGTTCCACACCTGGTTGACTGAAGAAGCTGACACTTTGAGTACCTTGAATATACCGTGTTC 1479

 CPRA 1453 GGGTTGGGTAACTCCACGTACGAGTTCTTCAATGCCATTGGTAGAAAGTTTGACAGATTGTTGAGCGAGA 1522
 CPRB 1480 GGGTTGGGTAACTCCACGTACGAGTTCTTCAATGCCATTGGTAGAAAGTTTGACAGATTGTTGAGTGA 1549

 CPRA 1523 AAGGTGGTGACAGGTTTGGCTGAATACGCTGAAGGTGATGACGGTACTGCCACCTTGGACGAAGATTTCAT 1592
 CPRB 1550 AAGGTGGTGACAGATTGCTGAATATGCTGAAGGTGACGACGGCACTGCCACCTTGGACGAAGATTTCAT 1619

 CPRA 1593 GGCCTGGAAGGACAAATGTCTTTGACGCCCTTGAAGAATGATTGAACCTTGAAGAAAGGAATTGAAGTAC 1662
 CPRB 1620 GGCCTGGAAGGATAATGTCTTTGACGCCCTTGAAGAATGACTGAACCTTGAAGAAAGGAATTGAAGTAC 1689

 CPRA 1653 GAACCAAACGTGAATTTGACTGAGAGAGACGACTTGTCTGCTGCTGACTCCCAAGTTTCTTTGGGTGAGC 1732
 CPRB 1690 GAACCAAACGTGAATTTGACTGAGAGAGATGACTTGTCTGCTGCTGACTCCCAAGTTTCTTTGGGTGAGC 1759

 CPRA 1733 CAAACAAAGATACATCAACTCCGAGGTCATCGACTTGACCAAGGGTCCATTGACCAACCCACCCATA 1802
 CPRB 1760 CAAACAAAGATACATCAACTCCGAGGTCATCGACTTGACCAAGGGTCCATTGACCAACCCACCCATA 1829

 CPRA 1803 CTTGGCCAGAAACACCGAGACGAGAGAGATTGTTGAGCTCCAGGACAGACACTGTATCCAGTTGAATTT 1872
 CPRB 1830 CTTGGCCAGGATCACCAGAGACCGAGAGAGATTGTTGAGCTCCAGGAGAGACACTGTATTCAGTTGAATTT 1899

 CPRA 1873 GACATTTCTGAATCGAACTTGAATATACACCACCGGTGACCATCTAGCTATCTGGCCATCCAACTCCGAGC 1942
 CPRB 1900 GACATTTCTGAATCGAACTTGAATATACACCACCGGTGACCATCTAGCTATCTGGCCATCCAACTCCGAGC 1969

 CPRA 1943 AAAACATTAAAGCAATTTGCCAAGTGTTCGGATTGGAAGATAAACTCGACACTGTTATTGAATTGAAGGC 2012
 CPRB 1970 AAAACATTAAAGCAATTTGCCAAGTGTTCGGATTGGAAGATAAACTCGACACTGTTATTGAATTGAAGGC 2039

 CPRA 2013 GTTGGACTCCACTTACACCATCCCATTTCCCAACCCCAATTACCTACGGTGGTGTCTATTAGACACCATTTA 2082
 CPRB 2040 GTTGGACTCCACTTACACCATCCCATTTCCCAACCCCAATTACCTACGGTGGTGTCTATTAGACACCATTTA 2109

 CPRA 2083 GAAATCTCCGGTCCAGTCTCGAGACAAATCTTTTGTCAATTGCTGGGTTTGGCTCTGATGAAGAAACAA 2152
 CPRB 2110 GAAATCTCCGGTCCAGTCTCGAGACAAATCTTTTGTCAATTGCTGGGTTTGGCTCTGATGAAGAAACAA 2179

 CPRA 2153 AGAAGGCTTTTACAGACTTGGTGGTGACAAAGCAATTCGCCGCCAAGGTACCCGCCAGAAAGTTCAA 2222
 CPRB 2180 AGAAGACTTTTACAGACTTGGTGGTGACAAAGCAATTCGCCGCCAAGGTACCCGCCAGAAAGTTCAA 2249

Figure 13B

CPRA 2223 CATTGCCGATGCCCTTGTATATATCTCCAAACAACGCTCCATGGTCCGATGTTCTTTTGAATTCCTTATT 2292
CPRB 2250 CATTGCCGATGCCCTTGTATATATCTCCAAACAACCTCCATGGTCCGATGTTCTTTTGAATTCCTTATT 2319
.....

CPRA 2293 GAAAAAGTTCCACACTTGACTCCACGTTACTACTCCATTTCTGCTTCCTCATTGAGTGAAAAAGCAACTCA 2362
CPRB 2320 GAAAAATCCAAACACTTGACTCCACGTTACTACTCCATTTCTTCTTCGTCGTTGAGTGAAAAAGCAACTCA 2389
.....

CPRA 2363 TCAACGTTACTGCAGTTGTTGAAGCCGAAGAAAGCTGATGGCAGACCACTCACTGGTGTGTGCACCA 2432
CPRB 2390 TCAATGTTACTGCAGTCGTTGAGGCCGAAGAAAGCCGATGGCAGACCACTCACTGGTGTGTGTACCA 2459
....

CPRA 2433 CTTGTTGAAGAACGTTGAAATTTGTGCAAAAACAAGCTGCCGAAAAAGCCACTTGTCCACTACGATTTGAGC 2502
CPRB 2460 CTTGTTGAAGAACATTGAAATTTGCGCAAAAACAAGCTGCCGAAAAAGCCACTTGTTCACTACGATTTGAGC 2529
.....

CPRA 2503 GGCCCAAGAGGCAAGTTCAACAAGTTCAAGTTGCCAGTGCAATGTGAGAAAGATCCAACTTTAAAGTTGCCA 2572
CPRB 2530 GGCCCAAGAGGCAAGTTCAACAAGTTCAAGTTGCCAGTGCAATGTGAGAAAGATCCAACTTTAAAGTTGCCA 2599
.....

CPRA 2573 AGAACTCCACCAACCCAGTTATCTTGATTGGTCCAGTACTGGTGTTCGCCCACTTGAGAGGTTTTGTCA 2642
CPRB 2600 AGAACTCCACCAACCCAGTTATCTTGATTGGTCCAGTACTGGTGTTCGCCCACTTGAGAGGTTTTGTCA 2669
.....

CPRA 2643 AGAAAGAGTTCAACAAGTCAAGAAATGGTGTCAATGTTGCCAAGACTTTGTTGTTTTATGGTTGCAGAAAC 2712
CPRB 2670 AGAAAGAGTTCAACAAGTCAAGAAATGGTGTCAATGTTGCCAAGACTTTGTTGTTTTATGGTTGCAGAAAC 2739
.....

CPRA 2713 TCCAAACGAGGACTTTTGTACAAAGCAAGAAATGGGCCGAGTACGCTTCTGTTTTGGGTGAAAACTTTGACA 2782
CPRB 2740 TCCAAACGAGGACTTTTGTACAAAGCAAGAAATGGGCCGAGTACGCTTCTGTTTTGGGTGAAAACTTTGACA 2809
.....

CPRA 2783 TGTTCATGCTTCTCCAGACAAGACCCATCCAAGAAAGTTTACGTCCAGGATAAGATTTTAGAAAAACAG 2852
CPRB 2810 TGTTCATGCTTCTCTAGACAAGACCCATCCAAGAAAGTTTACGTCCAGGATAAGATTTTAGAAAAACAG 2879
.....

CPRA 2853 CCAACTTGTGCACGAGTTGTTGACTGAAGGTGCCATTATCTACGTCTGTGGTGATGCCAGTAGAATGGCT 2922
CPRB 2880 CCAACTTGTGCACGAAATGTTGACCGAAGGTGCCATTATCTACGTCTGTGGTGACGCCAGTAGAATGGCC 2949
.....

CPRA 2923 AGAGACGTGCAGACCAATTTCCAAAGATTGTTGCTAAAAGCAGAGAAATAGTGAAAGCAAGGCTGCTG 2992
CPRB 2950 AGAGACGTGCAGACCAAGATCTCCAAAGATTGTTGCCAAAGCAGAGAAATCAGTGAAGCAAGGCCGCTG 3019
.....

CPRA 2993 AATTGGTCAAGTCTCGAAGGTCCTGAAAGTCCAAATAGATACCAAGAAAGATGTTTGGTAGACTCAAACGAATCTCTC 3062
CPRB 3020 AATTGGTCAAGTCTCGAAGTCCAAATAGATACCAAGAAAGATGTTTGGTAGACTCAAACGAATCTCTC 3089
.....

CPRA 3063 TTTCTCCCAACGCATTATGAAATCTTATCTCATTGAAGCTTTACATATGTTCTACATTTTATTTTTT 3132
CPRB 3090 TTTCTCCCAACGCATTATGAA-----TATTCTCATTGAAGTTTACATATGTTCTATATTTTATTTTTT 3155
.....

CPRA 3133 TTTTTTTTTTATTATTATATTACGAAACATAGGTCAACTATATATACTTGATTAAATGTTATAGAAACA 3202
CPRB 3155 TTT-----ATTATATTACGAAACATAGGTCAACTATATATACTTGATTAAATGTTATAGAAACA 3215
....

CPRA 3203 TAACTATTATCTACTCGTCTACTTCTTTGGCATTGACATCAACATTACCGTTCCCATTAACCGTTGCCGTT 3272
CPRB 3216 TAACTATTATCTACTCGTCTACTTCTTTGGCATTGCGATTGGCATTGGCATTGGCATTGGCATTGGCATT 3285
....

CPRA 3273 GGCAATGCCGGGATATTTAGTACAGTATCTCCAATCCGGATTGAGCTATTGTAGATCAGCTGCAAGTCA 3342
CPRB 3285 GGTAATGCCGGGATATTTAGTACAGTATCTCCAATCCGGATTGAGCTATTGTAAATCAGCTGCAAGTCA 3355
..

CPRA 3343 TTCTCCACCTTCACCAAGTACTTATACTTCATCTTTGACTTCAAGTCCAAAGTCATAAATATTACAAGTTA 3412
CPRB 3355 TTCTCCACCTTCACCAAGTACTTATACTTCATCTTTGACTTCAAGTCCAAAGTCATAAATATTACAAGTTA 3425
.....

Figure 13C

CPRA 3413 GCAAGAACTTCTGGCCATCCAGATATAGACGTTATTACGTTATTATGCGACGTATGGATGTGGTTATC 3482
CPRB 3426 GCAAGAACTTCTGGCCATCCACAAATATAGACGTTATTACGTTATTATGCGACGTATGGATATGGTTATC 3495
.....

CPRA 3483 CTTATTGAACCTTCTCAAACCTTCAAAAACAACCCACGTCGCCCAACGTCATTATCAACGACAAGTTCTGG 3552
CPRB 3496 CTTATTGAACCTTCTCAAACCTTCAAAAACAACCCACGTCGCCCAACGTCATTATCAACGACAAGTTCTGA 3565
.....

CPRA 3553 CTCACGTCGTCGGAGCTCGTCAAGTTCTCAATTAGATCGTTCTTGTATTGATCTTCTGGTACTTTCTCA 3622
CPRB 3566 CTCACGTCGTCGGAGCTCGTCAAGTTCTCAATTAGATCGTTCTTGTATTGATCTTCTGGTACTTTCTCA 3635
.....

CPRA 3623 ATTGCTGGAAACACATTGTCTCTCGTTGTTCAAATAGATCTTGAACAACTTTTCAACGGGATCAACTTCTC 3692
CPRB 3636 ACTGCTGGAAACACATTGTCTCTCGTTGTTCAAATAGATCTTGAACAACTTCTTCAAGGGAATCAACTTTTC 3705
*

CPRA 3693 AATCTGGGCCAAGATCTCCGCCGGGATCTTCAGAAACAAGTCTGCAACCCCTGGTCGATGGTCTCCGGG 3762
CPRB 3706 GATCTGGGCCAAGATTTCCGCCGGGATCTTCAGAAACAAGTCTGCAACCCCTGGTCGATGGTCTCCGGG 3775
.....

CPRA 3763 TACAACAAGTCCAAGGGGCAGAAGTGTCTAGGCACGTGTTTCACTGGTTCAACGAACATGTTCCGACAGT 3832
CPRB 3776 TACAACAAGTCTAAGGGGCAGAAGTGTCTAGGCACGTGTTTCACTGGTTCAAGGAACATGTTCCGACAGT 3845
.....

CPRA 3833 AGTTCGAGTTATAGTTATCGTACAACCAITTTGGTTTGATTTCGAAAATGACGGAGCTGATGCCATCATT 3902
CPRB 3846 AGTTCGAGTTATAGTTATCGTACAACCAITTTGGCTTGATTTCGAAAATGACGGAGCTGATCCCATCATT 3915
.....

CPRA 3903 CTCCTGGTTCTCTCATAGTACAACCTGGCACTTCTTCGAGAGGCTCAATTCTCTAGTTCCTCGTCCAAG 3972
CPRB 3916 CTCCTGGTTCTTTTCATAGTACAACCTGGCACTTCTTCGAGAGACTCACTCTCTAGTTCCTCGTCCAAG 3985
.....

CPRA 3973 ATATTCCGCAACAAGAGCCCGTACCGCTCAGGAGCATCAAGTCGTGGCCCTGGTTGTTCAACTTGTGA 4042
CPRB 3986 ATATTCCGCAACAAGAGCCCGTAGCGCTCAGGAGCATCAAGTCGTGGCCCTGGTTGTTCAACTTGTGA 4055
.....

CPRA 4043 TGAAGTCCGAGGTCAAGACAATCAACTGGATGTCGATGATCTGGTCCGGGAACAAGTTCTTGCACTTTAG 4112
CPRB 4056 TGAAGTCCGAGGTCAAGACAATCAACTGGATGTCGATGATCTGGTCCGGGAACAAGTTCTTGCACTTTAG 4125
.....

CPRA 4113 CTCGATGAAGTCGTACAACCTCACAGTCGAGATATCTCTGTTCTCTCTCAAGAGCCGGATCCGCAAG 4182
CPRB 4126 CTCGATGAAGTCGTACAACCT 4145
.....

CPRA 4183 AGCTTGTGCTTCAAGTAGTCGTTG 4205
CPRB 4146 4145

Figure 13D

CPRA	MALDKLDLYVIITLVVAVAAAYFAKNQFLDQPDGTGFLNDSGSNSRDVLLTLGGNNKNTL	60
CPRB	MALDKLDLYVIITLVVAVAAAYFAKNQFLDQPDGTGFLNDSGSNSRDVLLTLGGNNKNTL	60
CPRA	LLFGSQTGTAEADYANKLSRELHSRFGKLTWADFADYDWDNFGDITEDILVFFIVATYGE	120
CPRB	LLFGSQTGTAEADYANKLSRELHSRFGKLTWADFADYDWDNFGDITEDILVFFIVATYGE	120
*		
CPRA	GEPTDNADEFHTWLTEADTLSTLKYTVFGLGNSTYEFFNAIGRKGDRLLSEKGGDRFAE	180
CPRB	GEPTDNADEFHTWLTEADTLSTLRYTVFGLGNSTYEFFNAIGRKGDRLLSEKGGDRFAE	180
CPRA	YAEGLDGTGTLDEDFMAKONVFDALKQDINTEEKELKYEPNVKLTERDDLSAADSQVSL	240
CPRB	YAEGLDGTGTLDEDFMAKONVFDALKQDINTEEKELKYEPNVKLTERDDLSAADSQVSL	240
*		
CPRA	GEPNKKYINSEGIDLTGKQPFDHHPYLARITE TRELFSSKDRHCIHVEFDISESNLKYTT	300
CPRB	GEPNKKYINSEGIDLTGKQPFDHHPYLARITE TRELFSSKERHCIHVEFDISESNLKYTT	300
CPRA	GDHLAIWPSNSDENIKQFAKCFGLDGLDVTIELKALDSTYTIPFPTPIYGA VIRHLE	360
CPRB	GDHLAIWPSNSDENIKQFAKCFGLDGLDVTIELKALDSTYTIPFPTPIYGA VIRHLE	360
* * *		
CPRA	ISGFVSRQFTLSLAGFADDEETKQAFTRLGQDKQSFPAKVTRKGNIA DALLYSSNNAFW	420
CPRB	ISGFVSRQFTLSLAGFADDEETKKTFTRLGQDKQSFPAKVTRKGNIA DALLYSSNNTFW	420
**		
CPRA	SDVTFEFLIENVPHLTPRYYSISSSSSLSEKQLINVTAVVEAEEDADGRFVTGVVTNLLKN	480
CPRB	SDVTFEFLIENIQHLTPRYYSISSSSSLSEKQLINVTAVVEAEEDADGRFVTGVVTNLLKN	480
* *		
CPRA	VEIVQNKTEGKPLVHYDLSGPRGKGNKTKLPVAVRRSNFKLEKNSTTFVILIGPGTGVP	540
CPRB	IEIAQNKTEGKPLVHYDLSGPRGKGNKTKLPVAVRRSNFKLEKNSTTFVILIGPGTGVP	540
CPRA	LRGFVRE RVQVKNGVNVGKTLLFYGCRNSNEDFLYKQEWAEYASVLGENTEMNAFSRQ	600
CPRB	LRGFVRE RVQVKNGVNVGKTLLFYGCRNSNEDFLYKQEWAEYASVLGENTEMNAFSRQ	600
CPRA	DPSKKVIVQDKILENSQLVHELLTEGAIYVCGDASRMARDVQTTISKIVAKSREISEDK	660
CPRB	DPSKKVIVQDKILENSQLVHELLTEGAIYVCGDASRMARDVQTTISKIVAKSREISEDK	660
CPRA	AAELVKSKKVVQNRVQEDVH	680
CPRB	AAELVKSKKVVQNRVQEDVH	680

Figure 14

C. tropicalis 20336 CYP52 DNA Alignment of DS Sequence

CYP52A1A	1		0
CYP52A2A	1	GACCTGTGACGCTTCCGGTGTCTTGCCACCACTCCCAAGTTGACCGACGCCAAGTCATGTACCACTTT	70
CYP52A2B	1		0
CYP52A3A	1	GACATCATAT	11
CYP52A3B	1		0
CYP52A5A	1		0
CYP52A5B	1	TTACAATCATGG	12
CYP52A8A	1		0
CYP52A8B	1		0
CYP52D4A	1		0
CYP52A1A	1	CATATGCCGCTAATCTTCTTTTCTTTTATCACAGGAGAACTATCCACCCCACTTC	59
CYP52A2A	71	ATTTCGGTTACACTTCCAGATGGCTGGTACTGAAGAAGGTGTACCGAACCACAAAGTACTTTCTCCG	140
CYP52A2B	1		0
CYP52A3A	12	GACCCGGTTATTTGCCCTCAGGTTGCTTATTGAGCCGTAAAGTGCAGTAGAACTTTGCCCTTGGGTTTC	81
CYP52A3B	1		0
CYP52A5A	1	TGGAGTC	7
CYP52A5B	13	AGCTCGCTAGGAACCCAGATGTCTGGGAGAAGCTCCGCGAAGAGGTCAACACGAACTTTGGCATGGAGTC	82
CYP52A8A	1		0
CYP52A8B	1		0
CYP52D4A	1		0
CYP52A1A	60	GAAACACAATGACAACTCTGCGTAACTTGCAAACTCTGTCTGACTAATTGAAAACCTCCGACGAGTCA	129
CYP52A2A	141	CTTGTTCCGGTCAACCACTTCTGGTGTTCACCCCAATGAAGTACGCTCAACAATTGTCTGACAAGATCTC	210
CYP52A2B	1	GCTCAACAATTGTCTGACAAGATCTC	26
CYP52A3A	82	AAACTCTAGTATAAATGGTGATAACTGGTTGCACTCTTGCCATAGGCATGAAAATAGGCCGTTATAGTACT	151
CYP52A3B	1		0
CYP52A5A	8	GCCAGACTTGCTCACTTTGACTCCCTTCGAAACTCAAAGTACGTTCAAGCGGTGCTCAACGAACGCTC	77
CYP52A5B	83	GCCAGACTTGCTCACTTTTGACTCTCTTAGAAGCTCAAAGTACGTTCAAGCGGTGCTCAACGAACGCTT	152
CYP52A8A	1		0
CYP52A8B	1	AAACCGATACAGAAGAAGACAGTCAA	28
CYP52D4A	1		0
CYP52A1A	130	GACCTCCAGTCAAACGGACAGACAGACAAACACTTGGTGCGATGTTCATACCTACAGACATGTCAACGGG	199
CYP52A2A	211	GCAACACAAGGCTAACGCCCTGGTTGTTGAACACCGGTTGGGTTGGTTCTTCTGCTGCTAGAGGTGGTAAG	280
CYP52A2B	27	GCAACACAAGGCTAACGCCCTGGTTGTTGAACACTGGTTGGGTTGGTTCTTCTGCTGCTAGAGGTGGTAAG	96
CYP52A3A	152	ATATTTAATAAGCGTAGGAGTATAGGAATGCATATGACCGGTTTTTCTATATTTTAAAGATAATCTTAGT	221
CYP52A3B	1	CCTGCAGA	8
CYP52A5A	78	CGTATCTACCCGGGGGTACCAAGAAACATGAAGACAG--CTACGTGCAACACGACGTTGCCACGGCGAGG	145
CYP52A5B	153	CGTATCTACCCGGGGGTGCCACGAACATGAAGACAG--CTACGTGCAACACGACGTTGCCCGGTGGAGG	220
CYP52A8A	1		0
CYP52A8B	29	CAAGAACGTTAATGTCAAACAGGCGCCAAGAAGACCG--TTTGGCGGACTTGAAGAATGTGGCATTTGC	96
CYP52D4A	1		0
CYP52A1A	200	TGTTAGACGACGGTTTCTTGCAAGAC-AGGTGTTGGCATCTCGTACGATGGCAACTGCAGGAGGTGTCG	258
CYP52A2A	281	AGATGCTCATTGAAGTACACACGAGCCATTTTGGACGCTATCCACTCTGGTGAATTGTCCAGGTTGAAT	350
CYP52A2B	97	AGATGTTCAATTGAAGTACACACGAGCCATTTTGGACGCTATCCACTCTGGTGAATTGTCCAGGTTGAAT	156
CYP52A3A	222	AAATTTTGTATTCTCAGTAGGATTTTCATCAATTTTCGCAACCAATTTGCGGCAAAAATGATTCTTTTAC	291
CYP52A3B	9	ATTCCGGGCGCGTCCAGACAGTAGCAGTTATGCAAGCATGTGATTGTGGTTTTTGCACCTGTTTGCAC	78
CYP52A5A	146	AGGCA-AAGACGGGCAAGGAACCTATCT-TGGTGCAGAAAGGACAGTCCGTTGGGTTGATTACTATTGCCA	213
CYP52A5B	221	AGGCA-AAGACGGTAAAGGAACCTATTT-TGGTGCAGAAAGGACAGTCCGTTGGGTTGATTACTATTGCCA	288
CYP52A8A	1		0
CYP52A8B	97	CCATG-ATGTTTATGTTCTGGAGAGGT-TTTTCAAGGAATCGTCATCTCCGCCACCAAGAACCA	164
CYP52D4A	1		0

Figure 15A

CYP52A1A 269 ACTTCTCCTTTAGGCAATAGAAAAAGACTAAGAGAACAGCGTTTTTACAGGTTGCATTGGTTAATGTAGT 338
CYP52A2A 351 ACGAAACTTTCCAGTCTTCAACTTGAATGTCCCAACCTCCTGTCCAGGTGTCCCAAGTGAAATCTTAA 420
CYP52A2B 167 ACGAGACTTTCCAGTCTTCAACTTGAATGTCCCAACCTCCTGTCCAGGTGTCCCAAGTGAAATCTTAA 236
CYP52A3A 292 GTCAAAAGCTGA-ATAGTGCAGTTTAAAGCACCTAAATCACAATACAGCCTCTAGATACGACAGAGAA 360
CYP52A3B 79 GACAAATGATCG-ACAGT-CGATT--ACGTAATCCATATTATTAGAGGGGTAAATAAAATATAATGGCA 144
CYP52A5A 214 CGCAGACGGACCCAGAGTATTTTGGGGCCGACGCTGGTGAGTTTAAAGCCGAGAGATGGTTTGATTCA-- 281
CYP52A5B 289 CGCAGACGGACCCAGAGTATTTTGGGGCAGATGCTGGTGAGTTCAAAACGGAGAGATGGTTTGATTCA-- 356
CYP52A8A 1 0
CYP52A8B 165 GTTAAAGAGATCCATATTCAACCCACCGCAAGGTGACAAATGCTCAACAAACAGCAACAACA-- 232
CYP52D4A 1 0

CYP52A1A 339 ATTTTTTTAGTCCAGCATTCTGTGGTGTGCTCTGGTTTCTAGAATAGGAAATCAGGAGAAATGCAAA 408
CYP52A2A 421 CCCAACCAAGGCTGGACCGG-AAAGTGTGTGACTCTTCAACAAAGGAAATCAAGTCTTTGGCTGGTAAAGT 489
CYP52A2B 237 CCCAACCAAGGCTGGACCGG-AAAGTGTGTGACTCTTCAACAAAGGAAATCAAGTCTTTGGCTGGTAAAGT 304
CYP52A3A 361 GCTCTTTATGATCTGAAGAGCATTAGAAATAGCT---ACTATGAGCCACTATTGGGTATATATTAGGGA 427
CYP52A3B 145 GCC----AGAATTTCAACATTTTGCAACATGCAAAAGATGAGAACTCCAAACAGAAAAATAAAAA 210
CYP52A5A 282 AGCATGAAGAACTTGGGGTGTAAATCTTCCCTTCAATGCTGGGCCAGGACTTGTCTGGGGCAGCAGT 351
CYP52A5B 357 AGCATGAAGAACTTGGGGTGTAAATCTTCCCTTCAATGCTGGGGCCCGGACTTGTCTGGGGCAGCAGT 426
CYP52A8A 1 0
CYP52A8B 233 ACCCCCCAACAAACAGTGGAAATATGCCAGTCAA-CAAGAGTGGTGACAGACGAGGGAGAAAAACGCAAG 301
CYP52D4A 1 0

CYP52A1A 409 TTCAGATGGAAGAAACAAAGAGATAAAAAACAAAACAACTAGTCTTTCACCAATAGAAATGTTG---- 474
CYP52A2A 490 TTGCTGAAAC--TTCAAGACTATGCTGACCAAGCTACCGCTGA--AGTGAGAGCTGCAAGTCCAGAG 555
CYP52A2B 305 TTGCTGAAAC--TTCAAGACTATGCTGACCAAGCTACCGCTGA--AGTGAGAGCTGCAAGTCCAGAG 370
CYP52A3A 428 TTGGTGCAATTAAGTACGTACTAATAACAGAGAAATACTTAAACCAATTTCTGGGTATATCTAGTGG 497
CYP52A3B 211 ACTCCGACG--ACTCCGAACCAACAAACAACTGGGGCCGCCAG--AATTATTGA--TATT----- 267
CYP52A5A 352 ACACCTTTGATTGAAGCGAGCTACTTGTCTAGTCCGGTTGGCCAGACCTAC-CGGGCAATAGATTG---- 416
CYP52A5B 427 ACACCTTTGATTGAAGCGAGCTATTGCTAGTCAAGTTGGCGCAGACCTAC-CGGGTAAATCGATTG---- 491
CYP52A8A 1 0
CYP52A8B 302 CAACAGTGGTCTGATGCAAGATCAGCTACACCGCTTATCAGGAAAAAC-AGGAGCTCCACAC-- 365
CYP52D4A 1 GATGTGGTGCTTGAATTTCTCGAGACCATCTTGTGAGGTGCCATGAATCTGTACCTG---- 58

CYP52A1A 475 -ATGATATCATCCACTCGCTAAACGAATCATGTGGGTGATCTTCTCTTAGTTTGGTCTATCATAAAC 543
CYP52A2A 556 CTTAAAGATATTTATTCATATTTAGTTTGCCTAATTTCTCATACCCATC-ATCATTCAACATAT 624
CYP52A2B 371 CTTAAAGATATTTATTCATATTTAGTTTGCCTAATTTCTCATACCCATC-ATCATTCAACATAT 439
CYP52A3A 498 -TGAGGACCTTTTCTGAACATTGGGGTCAACCTTTTGTGAGTGCAGATCGATTTTCTGTTTGTGT 566
CYP52A3B 268 ---GTGACTTTTATTTATTTTCCGTTAA--CTTTCATTGCAAGTGAAGTGT--GTTACACGGGGTGGT 329
CYP52A5A 417 -CAGCCAGGATCGGGTACC-CACCAAGAAAGAACTCGTTGATCAACATAGTGTCTGCCGAGGGGTGTT 484
CYP52A5B 492 -CTGCCAGGGTGGGGTACC-CACCAAGAAAGAACTCGTTGATCAATATGAGTGTCTGCCGAGGGGTGGT 559
CYP52A8A 1 0
CYP52A8B 367 -CATATGCCCATCAGAGCAACACCAAGCAAGTTAGTGTATAGTGTCTGTAGTTAAGTCAATGCAATGTA 435
CYP52D4A 59 -TCTGTAAAGCACAGGGAATGCTTCAACCTTATGCAATTTCTGTCTATTGCAAGCGTGTGCTGCAC 127

CYP52A1A 544 ACATGAAGGTGAATCCAAA-TACACTACACTCCGGGTATTGTCTCTTCTTTACAGATGTCTCATTTGTC 612
CYP52A2A 625 ATATAAAGTTACTTCGGA-----TATCATTTGAATCGTGGGTGTGCGCAATTGGATGATTGGAA 683
CYP52A2B 440 ATATAAAGTTATTTCGGAAC-TCATA---TATCATTTGAATCGTGGGTGTGCGCAATTGGATGATTGGAA 505
CYP52A3A 567 AATAATAGTGAACCTTTGTG-TAATAAATCTTCATGCAAGACTTGCAATTTGAGCTTGGGAGTTCAAG 635
CYP52A3B 330 GATGGTGTGGTTTCTACAA-TGCAAGGGCAGAGTGAAGGTTTCCACATAACGT-TGCACCATATCAAC 397
CYP52A5A 485 TGT--AAAGCTTTATAAGGA-TGTAAAGGTAGATGATAGTTGTGTAGGAGGAGCGGAGATAAATTAGAT 551
CYP52A5B 560 TGT--AAAGTTTCAAGGA-TCTAGATGGATATGTA-AGGTGTGAGGAGGAGCGGAGATAAATTAGAT 625
CYP52A8A 1 0
CYP52A8B 436 CCA--ATAAGACTATCCCTT-CTTACAACCAAGTTTCTGCCGCCCTGTCTGGCA-ACAGATGCTGGCC 501
CYP52D4A 128 GATATCTGCCAAGGTATATAGCAGAACGTGCTAGTGTCTCTCCGGTCATATTCTGTGGTATTCTGCA 197

Figure 15B

CYP52A1A 269 ACTTCTCCTTTAGGCAATAGAAAAAGACTAAGAGAACAGCGTTTTTACAGGTTGCATTGGTTAATGTAGT .338
 CYP52A2A 351 ACGAAACTTTCCAGTCTTCAACTTGAATGTCCCAACCTCCTGTCCAGGTGTCCCAAGTGAATCTTTAA 420
 CYP52A2B 167 ACGAGACTTTCAGTCTTCAACTTGAATGTCCCAACCTCCTGTCCAGGTGTCCCAAGTGAATCTTTAA 236
 CYP52A3A 292 GTCAAAAGCTGA-ATAGTGCAGTTTAAAGCACCTAAAATCACAATACAGCCTCTAGATACGACAGAA 360
 CYP52A3B 79 GACAAATGATCG-ACAGT-CGATT--ACGTAATCCATATTTTATAGAGGGTAAATAAAAAATAATGGCA 144
 CYP52A5A 214 CGCAGACGGACCCAGAGTATTTGGGGCCGACGCTGGTGAAGTTAAGCCGGAGAGATGGTTTGATTCA-- 281
 CYP52A5B 289 CGCAGACGGACCCAGAGTATTTGGGGCAGATGCTGGTGAGTTCAAACCGGAGAGATGGTTTGATTCA-- 356
 CYP52A8A 1 0
 CYP52A8B 165 GTTAACGAGATCCATATTCAACCCACCGCAAGGTGACAAATGCTCAACAAACAGCAACAAACA-- 232
 CYP52D4A 1 0

CYP52A1A 339 ATTTTTTTAGTCCAGCATTCTGTGGGTTGCTCTGGGTTTCTAGAATAGGAAATCACAGGAGATGCAAA 408
 CYP52A2A 421 CCCAACCAAGGCTGGACCGG-AGGTGTTGACTCCTTCAACAAGGAATCAAGTCTTTGGCTGGTAAAGT 489
 CYP52A2B 237 CCCAACCAAGGCTGGACCGG-AGGTGTTGACTCCTTCAACAAGGAATCAAGTCTTTGGCTGGTAAAGT 304
 CYP52A3A 361 GCTCTTTATGATCTGAAGAAGCATTAGAATAGCT---ACTATGAGCCACTATTGGTGATATATTAGGGA 427
 CYP52A3B 145 GCC---AGAATTTCAAAACATTTTGCAAAAGATGAGAACTCACAAGGAAAAATAAAAA 210
 CYP52A5A 282 AGCATGAAGAAGCTTGGGGTGTAAATCTTGGCGTTCAATGCTGGGCCACGGACTTGGCTGGGGCAGCAGT 351
 CYP52A5B 357 AGCATGAAGAAGCTTGGGGTGTAAATCTTGGCGTTCAATGCTGGGCCACGGACTTGGCTGGGGCAGCAGT 426
 CYP52A8A 1 0
 CYP52A8B 233 ACCCCACAAAGAACAGTGGAAATAATGCCAGTCAA-CAAAGAGTGGTGACAGACGAGGGAGAAAAACGCAAG 301
 CYP52D4A 1 0

CYP52A1A 409 TTCAGATGGAAGAACAAGAGATAAAAAACAAGAACTGAGTTTGCACCAATAGAATGTTG---- 474
 CYP52A2A 490 TTGCTGAAGAAC--TTCAAGACCTATGCTGACCAAGCTACCGCTGA--AGTGAGAGCTGCAGGTCCAGAA 555
 CYP52A2B 305 TTGCTGAAGAAC--TTCAAGACCTATGCTGACCAAGCTACCGCTGA--AGTGAGAGCTGCAGGTCCAGAA 370
 CYP52A3A 428 TTGGTGCAATTAAGTACGTACTAATAAACAAGAGAAATACTTAACCAATTTCTGGTGATATCTTAGTGG 497
 CYP52A3B 211 ACTCCGCAAGC--ACTCCGAACCAACAAAACAATGGGGGGCGCCAG--AATTATTGAC--TATT----- 267
 CYP52A5A 352 ACACCTTTGATTGAAGCGAGCTACTTGTCTAGTCCGGTTGGCCACAGCTAC-CGGGCAATAGATTG---- 416
 CYP52A5B 427 ACACCTTTGATTGAAGCGAGCTATTGTCTAGTCAAGTTGGCGCAGACCTAC-CGGGTAATCGATTG---- 491
 CYP52A8A 1 0
 CYP52A8B 302 CAACAGTGGTTCTGATGCAAGATCAGCTACACCGCTTCATCAAGAAAGC-AGGAGCTCCCAACAC---- 366
 CYP52D4A 1 GATGTGGTGCTTGATTCTCGAGACACATCCTTGTAGGTGCCATGAATCTGTACTCTG---- 58

CYP52A1A 475 -ATGATATCATCCACTCGCTAAACGAATCATGTGGTGATCTTCTCTTTAGTTTTGGTCTATCATAAAC 543
 CYP52A2A 555 CTTAAAGATATTTATTCATTATTTAGTTTGGCTATTTATTTCTCATTAACCATC-ATCATTCAACACTAT 624
 CYP52A2B 371 CTTAAAGATATTTATTCATTATTTAGTTTGGCTATTTATTTCTCATTAACCATC-ATCATTCAACACTAT 439
 CYP52A3A 498 -TGAGGACCTTTTCTGAACATTCCGGTCAAACTTTTGTGAGTGCGACATCGATTTTTCGTTTGTGT 566
 CYP52A3B 268 ----GTGACTTTTATTTTATTTTCCGTTAA--CTTCATTGCAAGTGAAGTGT--GTTACACGGGGTGGT 329
 CYP52A5A 417 -CAGCCAGGATCGGCGTACC-CACCAAGAAAGAGTGGTGTATCAACATGAGTGTGCGCGACGGGGTGT 484
 CYP52A5B 492 -CTGCCAGGGTCCGCGTACC-CACCAAGAAAGAGTGGTGTATCAACATGAGTGTGCGCGACGGGGTGGT 559
 CYP52A8A 1 0
 CYP52A8B 367 -CATATGCCCATCAGGAGCAACACAGCAGGTTAGTGTATAGTAGTCTGTAGTTAAGTCAATGCAATGTA 435
 CYP52D4A 59 -TCTGTAAAGCAAGGAACTGCTTCAACACCTTATTGCATATTCTGTCTATTGCAAGCGTGTGCTGCAAC 127

CYP52A1A 544 ACATGAAGTGAATCCAAA-TACACTACACTCCGGGTATTGTCTTCTGTTTACAGATGTCTCATTGTC 612
 CYP52A2A 625 ATATAAAGTTACTTCGGA-----TATCAITGTAAATCGTGGGTGTCGCAATTGGATGATTGGAA 683
 CYP52A2B 440 ATATAAAGTTATTTCGGAAC-TCATA---TATCAITGTAAATCGTGGGTGTCGCAATTGGATGATTGGAA 505
 CYP52A3A 567 AATAATAGTGAACCTTTGTG-TAATAATCTTCATGCAAGACTTGCAATTCGAGCTTGGAGTTTCAGG 635
 CYP52A3B 330 GATGGTGTGTGTTTCTACAA-TGCAAGGGCAGCTTGAAGGTTTCCACATAACGT-TGCACCATATCAAC 397
 CYP52A5A 465 TGT--AAAGCTTTATAAGGA-TGTAAAGGTAGATGATAGTTGTGTAGGAGGAGCGGAGATAAATTAGAT 551
 CYP52A5B 560 TGT--AAAGCTTTACAAGGA-TCTAGATGGATATGTA-AGGTGTGTAGGAGGAGCGGAGATAAATTAGAT 625
 CYP52A8A 1 0
 CYP52A8B 436 CCA--ATAAGACTATCCCTT-CTTACAACCAAGTTTCTGCCGCGCTGTCTGGCA-ACAGATGCTGGCC 501
 CYP52D4A 128 GATATCTGCCAAGGTATATAGCAGAACGTGCTGATGGTTCTCCGGTCAATATTCTGTTGGTAGTTCTGCA 197

Figure 15B

CYP52A1A	613	TTACTTTTGAAGTTCATAGGAGTTGCCGTGTGAGAGATCACAGAGATTATCACACTCACATTTATCGTAGTT	682
CYP52A2A	684	CTGCGCTTGAACCGGATTATGTCACGAGCGGAGA-TAAAGATTACGT---AATTTATCTCTCTGA7ACA	749
CYP52A2B	506	CTGTAGTTGGAAACGGATTATGTCACGATGCGGAGA-TAACACG-----AGATTATCTCTCTAAG7CA	565
CYP52A3A	636	C---CAATTTGACCTCGTTATGTGATAAAAGAAAGCCAAAGGTAATT---AGCAGACGC---AATGGG	697
CYP52A3B	398	T---CAATTTATCTCTCATGTGATAAAAGAAAGCCAAAGGTAATT---GGCAGACCCCCAAGGGG	462
CYP52A5A	552	TTGATTTTG---TGTAAAGTTTTGGATGTCAACCTACTCCGCACITTCATGCA-GTGTGTGTGACACAAG	617
CYP52A5B	626	TTGATTTTG---TGTAAAGTTTAGCACGTCAAGCTACTCCGCACITTTGT---GTGTAGGAGCACAA---	685
CYP52A8A	1		0
CYP52A8B	502	GACACACTT---TCAACTGAGTTTGGTCTAGAATTCTTGACATGACAGACA-AGGAAACTCTTACAAAG	567
CYP52D4A	198	GGTAAATTTGGATGTGAGGTAGTGGAGGAGGTTTGTATCGGTTGTGTT-TTCTTCTTCTCTCTCTCTG	266
CYP52A1A	683	TCCTATCTCATGCTGTGTGTCTCTGGTTGGTTTCATGAGTTTGGATT--GTTGTACATTAAAGGAATCGCT	750
CYP52A2A	750	ATTTTAGCCGTGTTCAACGCCCTTCTTTGTT-CTGAGCGAAGGAT--AAATAATTAGACTTCCACAGCT	816
CYP52A2B	566	ATTTTGGCCTCATTACACGCCCTTCTT-----CTGAGCTAAGGAT--AAATAATTAGACTTCCACAGTT	629
CYP52A3A	698	AACATGGAGTGGAAAGCAATGGAAAGCAGCCCC-AGGACGGAGTAATTTAGTCCACACTACATCTGGGGT	766
CYP52A3B	463	AACACGGAGTAGAAAGCAATGGAAAGCAGCCCC-ATGACAGTGCCATTTAGCCCAACAACACATCTAGTATT	531
CYP52A5A	618	GTGTACTACGTGTGCGTGTGCGCCAGAGACA---GCCCAAGGGGG--TGGTAGTGT-GTGTGTGCGGAA	681
CYP52A5B	686	---TACTCCGTCTGCGCTGTGCGCCAGAGACA---GCCCAAGGGG-----TAGTGT-GTGTGTGCGGAA	741
CYP52A8A	1	GAATTCCTTTGGATCTAATTCACGCTGATC---TTGCTAATCCT--TATCAACGTAGTTGTGATCATT	62
CYP52A8B	568	--ACAACACTGTGCTCTGATGCCACITGATC---TTGCTAAGCCT--TATCAACGTAAATTGAGATCATT	630
CYP52D4A	267	ATTCACCTCCAGCTCTCTTGGGTTCTGTGCTGTGCTGAGTC--GTACTGTTGGATTAAAGTCCATC	334
CYP52A1A	751	GGAAAGCAAAGCTAACTAAATTTCTTTGTACAGGTACACTAACCTGTAAAGCTTCACTGCCACGCCAG	820
CYP52A2A	817	CATTCTAATTTCCGT---CACGCGAATATTGAA-----GGGGGTACATGTGGCCGCTGAA--	869
CYP52A2B	629	CATTAAATATCCGT---CACGCGAATATTGAA-----GGGGGTACATGTGGCCGCTGAA--	694
CYP52A3A	767	---TTTTTTTTTGTGCGCAAGTACACACCTGGACT-TTAGTTTTTGCCTCATAAAGTTAAACAATCTAA-	830
CYP52A3B	532	CTTTTTTTTTTTTGTGCGCAAGTACACACCTGGACT-TTAGTTTTTGCCTCATAAAGTTAAACAATCTCA-	599
CYP52A5A	682	GTGCATGTGACACA---ACGCGTGGGTTCTGCGCAATGGTGGACTAAGTGACAGGTAAGCAGCGACCTGA	748
CYP52A5B	742	GTGCATGTGACACA---ATACCTTGGTTCTGCGCAATGGGATTTAGTGTAGGTAAGCTGCGACCTGA	808
CYP52A8A	63	GTTTGTCTGAATTAT--ACACACAGTGGAAAGAAATATGGTCTAATTTGACAGTCCCACTGGCATTGTG--	128
CYP52A8B	631	GTTTGTCTGAATTAT--ACACACAGTGGAAAGAAATATGGTCTAATTTGACAGTCCCACTGGCATTGTG--	696
CYP52D4A	335	GCATGTGTGAAAAAAGTAGCGCTTATTAGACAAACAGTTGTTGGGCGGGTATCAGAAATAGTCTGTT	404
CYP52A1A	821	TCCTTCTGATTGGGCAAGTGACAACTACA-ACCTGCAAAACAG---CACTCCGCTTGTACAGGTT	885
CYP52A2A	870	-TGTGGGGG--CAGTAAACGCACTCTCTC-----CTCTCCAGGAATAGTGCAACGG	918
CYP52A2B	695	-TGTGGGGTGCAGTAAACGCACTCTCTCTCTCCCCCCCCCCCCCCCCCTCAGGAATAGTACAACGG	763
CYP52A3A	831	-CCTTTGGC-TCTCCACTCTCTCGCGCCCAATATTCGTTTTT-ACACCTCAGGCTAGCGACAGCAC	897
CYP52A3B	600	-CCTTTGGC-TCTCCAGTGTCTCGCGCTCCAGATGCTCGTTTT-ACACCTCAGGCTAAGCAACAACAC	665
CYP52A5A	749	ACATTCCTCAACGCTTAAAGACACTGGTGG-TAGAGATGCGGACCAAG---CTATTCTTGTCTGT-GCTA	811
CYP52A5B	809	ACATTCCTCAACGCTTAAAGACACTGGTGG-TAGAGATGCGGACCAAG---GGCTATTCTTGTCTGT-GCTA	875
CYP52A8A	129	-TGTTT-----GTGGGGGGGGGGGGTGCACACATTTTATGTCGA---TTCTTTGTGATTAC-CCCT	187
CYP52A8B	697	-TGTTT--GGGGGGGGGGGGGGGGTGCACACATTTTATGTCGAATGTTTGTGTTGCTGTTCC-CCCT	762
CYP52D4A	405	GTGACGACCAATGATGCAACTTGACGAGAGCTCGTTAGGA-----ATCCACAGAAATGATAGCAGGA	469
CYP52A1A	885	GTCTCTCTCAACCAACAAAAAATAGATTAACTTTCTTTGCTCATGCAATCGGAGTTATCTCTG	955
CYP52A2A	919	AGGAAGGATAACGGATAGAAAGCGGAATGCGAGGAAT--TTGAAACGCGCAAGAAAGCAATATCCGG	986
CYP52A2B	764	GGGAAGGATAACGGATAGCAAGTGGAAATGCGAGGAAT--TTGAAACGCGCAAGAAAGCAATATCCGG	831
CYP52A3A	898	AAACCCCATAGAGGAATGGGCAAGTTAAACACTTTTGGCTTCAATGATTCTATTCTGCTACTACATT	967
CYP52A3B	665	AAACCCCATAGAGGAATGGG-CAAGTTAAACACTTTTGGTTCAATGATTCTATTCTGCTACTACT-----	729
CYP52A5A	812	CCCCGGCATGGA-AAATCACTGCGGGAAGAA--TAAATTTATCCGTAGAAATCCACAGAGCG-----G	872
CYP52A5B	876	CCCC-TGCACGGA-AAATCGATTGAGGGAAGAA--CAATTTATCCGTAGAAATCCACAGAGCG-----G	935
CYP52A8A	188	CCCCCTATCAT---TCATTCCCAAGGATTAG--TTTTTCTCACTGGAAATTCGCTGTCC-----	244
CYP52A8B	763	CCCCCTCCCCCTATCATGCCCAAGGATTAG--TTTTTCTCACTGGAAATTCGCTGTCC-----	822
CYP52D4A	470	GCTTACTACGTGAGAGATTCTGCTTAGAGGATG--TTCTTCTTGTGATTCCATTAGGTGGGTATCAT	537

Figure 15C

CYP52A1A	956	A--AAGAGTGGCCTTTGTGTAATGTGTGCCAAA-CTCAAACTCGCAAACTAACCCACAGAATGAT-----	1016
CYP52A2A	987	GCTACCAGGTTTTTGAGCCAGGGGAACAACACTCCTATTCTCGCTCAATGACTGAACATAGAAAAA----	1050
CYP52A2B	832	GCTATCACGTTTTTGAGCCAGGGGAACAACACTCCT-CTTCGCAAAAAAATTAACGTAGACAAAAA--A	900
CYP52A3A	968	CTTCTCTGTGTTTTGTGCTTTGAAATGCACCATGTGAAATAAACGCAAAATTATATATACCTTTTCATC-	1034
CYP52A3B	730	----CTCTGTGTTTTGTGTTTTGATTGACCATGTGAAATAAACGCAAAATTATATATACCTTTTCGTG---	793
CYP52A5A	873	A--TAAATTTGCCCACTCCTCATCATCAACGAG-CGGCCACTAACTACATCATCTCCCCTATTIT-----	933
CYP52A5B	936	A--TAAATTTGTGCATTTGCTGGCTTGGCCAC-----CCACGACATTCT-----	978
CYP52A8A	245	-----ACCTGTCAACCCCCCCCCCCCCCCCC-CCA CTGGC--CTACCTCGCCCTGC-----	293
CYP52A8B	823	-----ACCTGTCAACCCCCCTCAC-----TGGCCTGCCCTGC-----	853
CYP52D4A	538	CTCGGTTGGTGACAACCTTGACACAAAGCAGTTCGAGAAACACCCACAACAATCACCATTCCAGC-----	606
CYP52A1A	1017	TTCCTCTCACAATTATATAAACTCACCCACATTTCCACAGACCGTAATTTTCATGTCTCAC-TTICTCTTTT	1085
CYP52A2A	1051	-----CACCAAGACGCAATGAAACGACATGGACATTTAGACCTCCCCACATGTGTATAGTTTGTCTTAAC	1115
CYP52A2B	901	AACTCCACCAAGACACAATGAATCGACATGGACATTTAGACCTCCCCACATGTGTGAAAGCTTCTCTGGCG	970
CYP52A3A	1035	CCTCCTCCTATATCTCTTTTGTGTAG-AATTGTGTTTTTATCGTTTCTGCTTTTGTGCACTCTCCCACTCC	1103
CYP52A3B	794	TGTCCTCCAAATGTCTCTTTTGTGTCGCAATTTGCTCTTTTGTCTTTTGTGCTCTCTCTCCCACTCC	863
CYP52A5A	934	CTCTCTCTCTCTTTGTCTTACTCCGCTCCCGTTTCCTTAGCCACAGATACACACCACT-GCAACACGCA	1002
CYP52A5B	979	TTTCTCTCTCTTTGTCTTACTCCGCTCCGTTTCCTTAGCCAAATTAACACCACTCATATAAAGAT	1048
CYP52A8A	294	CCTGCACGTCCTGTGTTTTGTGCTGTCTTTCCACGCTATAAAGCCCTGGCGTCGGGCGAAGGTTTT	363
CYP52A8B	854	CTGCACGCTCTGTGTTTTGTGCTGTGCGCACTCCACGCTATAAAGCCCTGGCGTACGGCCAAAGTTTT	923
CYP52D4A	602	TATCACTTCTACATGTCAACCTACGATGTATCTCATCACTATAGTTTCTTGGCAATCGTTTATTTGTT	671
CYP52A1A	1086	GCTCTTCTTTTACTTAGTCAGGTTTTGATATGCTTCTTTTTTATTACCTATCTTATTATTATTATTATTC	1155
CYP52A2A	1116	AGA-----AAAGTATAATAAGAAACCAATGACCGTCCCTTTTTCTTGGCGCTTCACTTTTTTTTTTTA	1179
CYP52A2B	971	AAAGCAAAAAAAGTATAATAAGGACCCATGCTTCCCTCTTCTGGGCGGTTTCACTTTTCTCTTTTCT	1040
CYP52A3A	1104	ACAA-----AGAAAAAATACTACATATATGTCGCTCTCTCCATCGTTT	1146
CYP52A3B	864	ACAATCAGTGCAGCAACACACAAAGAAAGAAAAAACCCTACATATGTCGCTCTCTCCATCGTTT	933
CYP52A5A	1003	GCA--ACAATATTAAGAGATACGCC-----AGGCCCACTTCTTTTCTTTTCTTCACTTTTTTGACTGC-A	1064
CYP52A5B	1049	ACG--CTAGCCCACTGTCTTTCT-----TTTCTTCACTTTTTTGGTGTTGCTTTTTTGGCGTGC-T	1110
CYP52A8A	364	TCCACCCAGCAAAAAAAGAGTCTAAAAAATTTGGTTGATCTTTTTGGTGTCAGAGGTTTT--CAC-C	429
CYP52A8B	924	CTCTCAGGCAAAAAA-----AATTGGCTGATCTTTTTGGGCTGCAAGGTTTTTACCAC-C	982
CYP52D4A	672	ATGGGTCACATCCAATACACTCCACCA--TGAGAAAGAAACGGAAGCAGAATACCAGAATGACA	739
CYP52A1A	1156	ATTTATACCAACCAACC--AAACATGGCCACACAGAAATCATGATTCGTACTTCCGTACTTGACCAA	1223
CYP52A2A	1180	TCTT-----ACACACTCATCGACCA--TGACTGTACACGATATTATCGCCACATCTTACCAA	1236
CYP52A2B	1041	TGTCTATCAACACACACACCTCAAGGACA--TGACTGTACAGGATATTATCGCCACATACATACCAA	1109
CYP52A3A	1147	GGCC-----AAGAGGGTTCTCGCTACCCTAGTCTTACATCGAGTACTTCTTGACA-ACTACACCAAG	1208
CYP52A3B	934	GGTC-----AGGAGGGTTCTCGCTACCCTAGTCTTACATCGAGTACTTCTTGACA-ACTACACCAAG	955
CYP52A5A	1065	ACTTTCTCAATCCACCAAGGCAACACACAGCGCTATGATGTGAACAACTCTTGAAATTT-----	1127
CYP52A5B	1111	ACTTTCTCAACC-----ACCAACCAACCAACCACTGATTTGAACAAATCCTAGATATT-----	1165
CYP52A8A	430	ACCACCTTCAACA--CCTCACTATTGGAACAA--AGATGGCTCGATCGATCTTACATTACT-----	488
CYP52A8B	593	ACCACCAACCA--CCTCACTATTCAACAA--AGATGGCTCGACCAAGATCTTACATTACT-----	1041
CYP52D4A	740	GTGTG----AGTTCTGACCAATTGCTAACTA--TGACTATATCTAGTTTGTCTCGTGGGATG-----	797
CYP52A1A	1224	ATGGTACACTGTGATTACTGCAGCAGTATTAGTCTTCTTATCTCCACAAACATCAAGAACTACGTCAAG	1293
CYP52A2A	1237	ATGGTACGCTGATAGTACCCTCGCTTGTGATTGCTTATAGAGTCTCGACTACTTCTATGGCAGATACTTG	1306
CYP52A2B	1110	ATGGTACGCTGATAGTACCCTCGCTTGTGATTGCTTATAGGCTCTCGACTACTTCTATGGCAGATACTTG	1179
CYP52A3A	1209	ATGGTACTACTTCACTCTTGGTGGCTTCTGCTTGAACCTTATTAAGATTGTGCTCCACAAAGGTACTTG	1278
CYP52A3B	996	ATGGTACTACTTCACTCTTGGTGGCTTCTGCTTGAACCTTATTAAGATTGTGCTCCACAAAGGTACTTG	1065
CYP52A5A	1167	--GGTATGCTGTTGGCAGTGTTGTACATCATCAACCACTCTTGTGATTAAGAAAGTCTGGCTCTTG	1234
CYP52A5B	1167	--GGTATGCTGTTGGCAGTGTTGTACATCATCAACCACTCTTGTGATTAAGAAAGTCTGGCTCTTG	1234
CYP52A8A	489	--GGTACATTTGCTTGGCAATGTTGGCCATTAACACAGATCGTGGCTCATGTCAGGACCAATTTATTT	556
CYP52A8B	1042	--GGTACATTTGCTTGGCAATGTTGGCCATTAACACAGATCGTGGCTCATGTCAGGACCAATTTATTT	1109
CYP52D4A	798	-TGATCTGTGCTGCTCTTCAATTTCGGTTTGTGTTTATTCGGGAT-GAATTTGTTATACTAAATACTTG	865

Figure 15D

[illegible]

Figure 15F

CYP52A1A	1975	GTCCACCACTTGGCCAAAGTACTTTGTCAA	CAAGGCGTTGAACTTTACTCTGAA	GAAGCTGGAAGAGAAAT	2044
CYP52A2A	1997	GTGCACAAAGTTTGCTGACTACTACGTCAA	CAAGGCGTTTGGACITGACGGCTGAA	CAATTTGGAAAGATGAGG	2066
CYP52A2B	1870	GTGCACAAAGTTTGCTGACTACTTACGTCA	GAGGCGTTTGGAATTCACACCTGAA	CAATTTGGAAAGAGCGAG	1939
CYP52A3A	1978	GTGCACAAAGTTTGCTGACCACTATGTGCAA	AAAGGCGTTTGGAGTTGACCGACGAT	GACTTTGCGAAGAACAG	2047
CYP52A3B	1765	GTGCACAAAGTTTGCTGACCACTATGTGCAA	AAAGGCGTTTGGAGTTGACCGACGAT	GACTTTGCGAAGAACAG	1834
CYP52A5A	1886	GTGCACAAAGTTTACCAACTACTATGTTCA	GAAAGCGTTTGGATGCTAGCCCAAGAA	GCTTTGAAAGAGCAA	1955
CYP52A5B	1923	GTGCACAAAGTTTACCAACTACTATGTTCA	GAAAGCGTTTGGATGCTAGCCCAAGAA	GCTTTGAAAGAGCAA	1994
CYP52A8A	1232	GTGCACAAAGTTTACCAACTACTATGTTCA	GAAAGCGTTTGGATGCTAGCCCAAGAA	GCTTTGAAAGAGCAA	1301
CYP52A8B	1785	GTGCACAAAGTTTACCAACTACTATGTTCA	GAAAGCGTTTGGATGCTAGCCCAAGAA	GCTTTGAAAGAGCAA	1854
CYP52D4A	1517	GTGCCAAAGTTTGCAGCCAGTGTGTC	CAAGGCGTTTACATGTTGCACCGGAA	GACACCC-----A	1577

CY752A1A	2045	CCAAAGTCCGGTTACGTTTTCCTGTACGAALTGGTTAAAGCAAACAGAGATCCAAAGGCTCTTGCAAGATCA	2114
CY752A2A	2067	ATGGTT-----ATGTGTTTTCGTACGAALTGGTCAAGCAAACAGAGATCCAAAGGCTGTTGAGAGACCA	2130
CY752A2B	1940	ATGGTT-----ATGTGTCCTTGTACGAAGTTGGTCAAGCAAACAGAGACAGCGCAATGTTGTGAGAGACCA	2003
CY752A3A	2048	ACGGCT-----ATGTGTTCTTGTACGAAGTTGGCTTAAGCAAACAGAGACCCCAAGGCTCTGAGAGACCA	2121
CY752A3B	1835	ACGGCT-----ATGTGTCCTTGTACGAAGTTGGCTTAAGCAAACAGAGACCCCAAGGCTCTGAGAGACCA	1898
CY752A3C	1956	GTTGGT-----ATGTGTCCTTGTACGAAGCTGTGCAAGCAACAGAGACCCCAATGTGTTGGCTGACCA	2019
CY752A3A	1956	GTTGGT-----ATGTGTCCTTGTATGAGCTTGTCAAGCAAGCAGAGACCCCAAGGTTGTCGGTGAACA	2058
CY752A3B	1955	GCGGGT-----ATGTGTTCTTGTATGAGCTTGTCAAGCAAGCAGAGACCCCAAGGTTGTCGGTGAACA	1365
CY752A8A	1302	GCGGGT-----ATGTGTTCTTGTATGAGCTTGTCAAGCAAGCAGAGACCCCAAGGTTGTCGGTGAACA	1918
CY752A9B	1855	GCGGGT-----ATGTGTCCTTGTACGAAGCTTGCACAGACAGCAAGAGACCCCAATGTGTTGCGTGAACA	1918
CY752D4A	1578	GCGAGT-----ACGGTCTTCTCCGGCAAGTTGGTCAAAACACAGTCGAGATCCCGTGTGTTTACAAAGACCA	1641

CY752A1A	2115	ATTGTTGAAACATTATGGTTGCCGGAAAGAGACACCACCTGCCGGTTTGTGTGCTTTGCTTTGCTTTGAATTG	2184
CY752A2A	2131	ATTGTTGAAACATCATGGTTGCTGGTAGAGACACACCGCCGGTTGTGTGCTTTGTTTCTTTGAATTG	2200
CY752A3A	2004	GTTGTTGAAACATCATGTTTCCGGTAGAGACACACCGCCGGTTGTGTGCTTTGTTTCTTTGAATTG	2073
CY752A3A	2112	GTTATGTGAACATTTTGGTTTCCGGTAGAGACACGACCGCCGGTTGTGTGCTTTGTTTCTTACAGAGTTG	2181
CY752A3A	1899	GTTGTTGAAACATTTTGGTTTCCGGTAGAGACACGACCGCCGGTTGTGTGCTTTGTTGTTACAGAGTTG	1968
CY752A5A	2020	GTCTTTGAACATCTGTTGGCCGGAAAGAGACACCACTGCTGGGTTGTGTGCTTTGCTGTCTTTGAGTTG	2089
CY752A5A	2059	GTCTTTGAACATCTGTTGGCCAGGAAAGAGACACCACTGCTGGGTTGTGTGCTTTGCTGTGTTGAGTTG	2128
CY752A8A	1366	GTCTTTGAACATCTGTTGGCCAGGAAAGAGACACCACTGCTGGGTTGTGTGCTTTGCTGTGTTGAGTTG	1435
CY752A8B	1919	GTCTTTGAACATCTGTTGGCTGAGAAAGGACACCACTGCTGGGTTGTGTGCTTTGCTGTGTTGAGTTG	1988
CY752D4A	1642	AGCGTTGAACGTTGCTGTTCTGGACGCGACACCAACCGCGTCGTTATATCGTTTGCAACATTTGAGCTA	1711

CY752A1A	2185	GCTAGACACCCAGAGATGTGGTCCAAAGTTGAGAGAAAGAAATCGAAAGTTAACTTTGGTGTTGGTGAAGCT	2254
CY752A2A	2201	GCCAGAAACCCAGAGATGTATCAACAAAGTTGAGAGAAAGAAATCGAGGACAAAGTTTGGACTCGGTTGAGAAATG	2270
CY752A2B	2074	GCCAGAAACCCAGAGAGGTGACCAACAAAGTTGAGAGAAAGAAATCGAGGACAAAGTTGGTCTTGGTGAGAAATG	2143
CY752A3A	2182	TCAGAGAAACCCGTAGAGGTTTGTCTAAGTTTGAGAGAGGAGGTTGGAAAAAGATTTGGACTCGGTTGAAGAGG	2251
CY752A3B	1969	TCAGAGAAACCCGTAGAGTGTTCGCCAAGTTTGAGAGAGGAGGTTGGAAAAAGATTTGGACTCGGCGAGAGAGG	2038
CY752A5A	2090	GCCAGACACCCAGAGATGTGGGCCAAAGTTGAGAGAGGAAATTCGAAACACAGATTTGGTCTCTAGAGAAAGACT	2159
CY752A5B	2129	GCCAGAAACCCACACATCTGGGCCAAAGTTGAGAGAGGAAATTCGAAACACAGATTTGGTCTCTGAGAGAAAGACT	2198
CY752A5C	1436	GCCAGAAACCCACACATCTGGGCCAAAGTTGAGAGAGGAAATTCGAAACACAGATTTGGTCTCTGGAGAAAGACT	1505
CY752A5D	1939	GCCAGGAAACCCACACATCTGGGCCAAAGTTGAGAGAGGAAATTCGAATCACAGATTTGGGCTCGGTGAGGACT	2058
CY752D4A	1712	GCCCCGAAATGACACATGTGGAGGAAAGCTACGAGAGGAGGTT-----ATCCTGA---CGATGGGACCG	1771

CY752A1A	2255	CCGCGGTGGAAGAAATTACCTTCGAAAGCCTTGAAGAGATGTGAATACTTGAAGGCTATCCTTAAACGAAAC	2324
CY752A2A	2271	CTAGTGTGTAAGACATTTCCTTTCGAGTCTGTTGAAGTCTCTGTGAATACTTGAAGGCTGTTCTTCAACGAAAC	2340
CY752A2B	2144	CTCGTGTTGAAGACACTTCCTTTCGAGTCTGTTGAAGTCATGTGAATACTTGAAGGCTGTTCTCAACGAAAC	2213
CY752A3A	2252	CTCGTGTTGAAGAGATCTCGTTTCGAGTCTGTTGAAGTCTGTTGAAGCTTGAAGGCTGTCTCAATGAAGAC	2321
CY752A3B	2039	CTCGTGTTGAAGAGATCTCTTTCGAGTCTGTTGAAGTCTGTGAGTACTTGAAGGCTGTCTCAATGAAGAC	2108
CY752A5A	2160	CTCGTGTTGAAGAGATTACCTTTCGAGAGCTTGAAGAGATGTGAGTACTTGAAGGCTTCTTCAATGAAGAC	2229
CY752A5B	2199	CTCGTGTTGAAGAGATTACCTTTCGAGAGCTTGAAGAGATGTGAGTACTTGAAGGCTTCTTCAACGAAAC	2268
CY752A8A	1506	CTCGTGTTGAAGAGATTAACCTTTGAGAGCTTGAAGAGATGTGAGTACTTGAAGGCGGTGTTGAACGAAAC	1575
CY752A9B	2059	CTCGTGTTGAAGAGATTACCTTTCGAGAGCTTGAAGAGATGTGAGTACTTGAAGGCGGTGTTGAACGAAAC	2128
CY752D4A	1772	TCCAG--TGATGAATAACCGTGGCGGGTGAAGAGTTCGGTATTACCTCAAGCAATCTCTAAACGAAAC	1839

Figure 15G

```

CYP52A1A 2325 CTTGCGTATGTACCCATCTGTTCTCTCACTTTAGAACCCGCCACGAGACACCACCTTTGCCAAGAGUT 2394
CYP52A2A 2341 CTTGAGATTGTACCCATCCGTGCCACAGAAITTCAGAGTTGCCACCAAGAACACTACCCCTCCCAAGTGT 2410
CYP52A2B 2214 TTTGAGATTGTACCCATCCGTGCCACAGAAITTCAGAGTTGCCACCAAGAACACTACCCCTCCCAAGAGGT 2283
CYP52A3A 2322 CTTGAGATTGTACCCATCCGTGCCACAGAAITTCAGAGTTGCCACCAAGAACACTACCCCTCCCAAGAGGT 2391
CYP52A3B 2109 CTTGAGATTGTACCCATCTGTTCTCAACAACTTCAGAGTTGCCACCAAGAACACTACCCCTCCCAAGAGGT 2178
CYP52A3A 2230 CTTGCGTATTTACCCAAAGTGTCCCAAGAACTTCAGAAITGCCACCAAGAACACTACCCCTCCCAAGAGGT 2299
CYP52A3B 2269 CTTGCGTATTTACCCAAAGTGTCCCAAGAACTTCAGAAITGCCACCAAGAACACTACCCCTCCCAAGAGGT 2338
CYP52A3A 1576 TTTGAGATTACACCCAAAGTGTCCCAAGAACTTCAGAAITGCCACCAAGAACACTACCCCTCCCAAGAGGT 1645
CYP52A3B 2129 GTTGAGATTACACCCAAAGTGTCCCAAGAACTTCAGAAITGCCACCAAGAACACTACCCCTCCCAAGAGGT 2198
CYP52D4A 1840 TCTTCGACTATACCCAAAGTGTCCCAAGAACTTCAGAAITGCCACCAAGAACACTACCCCTCCCAAGAGGT 1909
* * * * *

CYP52A1A 2395 GGTGGTGTAAAGGTACCGGACCCCAATCTACATTCTTAAAGGCTCCACTGTTGCTTACGTTGTCTACAAAG 2464
CYP52A2A 2411 GGTGGTAAAGGACGGGTGTCTCTCTGTTTGTGAGAAAGGGTCAAGCCGTTATTACCGGTGTCTACCGCA 2480
CYP52A2B 2284 GGTGGTAAAGGACGGGTATCTCTCTGTTTGTGAGAAAGGGTCAAGCCGTTATTACCGGTGTCTACCGCTG 2353
CYP52A3A 2392 GGTGGTGAAGATGGATCTCGCCCAATTTGCTGTCAAGAAAGGGTCAAGTTGTCTATGTACACTGTCTATGTCTA 2461
CYP52A3B 2179 GGTGGTAAAGACGGATGCTCGCCCAATTTGCTGTCAAGAAAGGGTCAAGTTGTCTATGTACACTGTCTATGTCTA 2248
CYP52A3A 2300 GGTGGTTCAGACGGTACCTCCCAATCTTGTATCCAAAGGGAGAAAGGTGTGTCTGTATGGTATCAACTCTA 2369
CYP52A3B 2339 GGTGGTTCAGACGGTACCCAGCCCAATCTTGTATCCAAAGGGAGAAAGGTGTGTCTGTATGGTATCAACTCTA 2408
CYP52A3A 1646 GGTGGCCCCAAGCGGCAAGGATCTATCTGTATCAAGAAAGGATGAGGTGGTGCAGTACTCCATCTCGGCCA 1715
CYP52A3B 2139 GGTGGCCCCAAGCGGCAAGGATCTATCTGTATCAAGAAAGGATGAGGTGGTGCAGTACTCCATCTCGGCCA 2268
CYP52D4A 1910 GGAGGTCCAGATGGATCGTTTCCGATTTTGATTAAGAAAGGGCCAGCCAGTGGGGTATTTCATTGTGTCTA 1979
* * * * *

CYP52A1A 2465 CCCACCGTTTGGAAAGATATCTACGGTAAAGGACGCTTAAAGGCTTCAGACCAAGAAAGATGGTTTGAAGCCATC 2534
CYP52A2A 2481 CCCACAGAAACCAGCTGTTTACGGTAAAGGACGCTTAAAGGCTTCAGACCAAGAAAGATGGTTTGAAGCCATC 2550
CYP52A2B 2354 CCCACAGAAACCAGCTGTTTACGGTAAAGGACGCTTAAAGGCTTCAGACCAAGAAAGATGGTTTGAAGCCATC 2423
CYP52A3A 2462 CCCACAGAGACCCAAATCTACGGTAAAGGACGCTTAAAGGCTTCAGACCAAGAAAGATGGTTTGAAGCCATC 2531
CYP52A3B 2249 CCCACAGAGACCCAAATCTACGGTAAAGGACGCTTAAAGGCTTCAGACCAAGAAAGATGGTTTGAAGCCATC 2318
CYP52A3A 2370 CTCATTTCGACCCCTGTCTATTATGGGCTGTATGCTGTCTGATTCAGACCAAGAAAGATGGTTTGAAGCCATC 2439
CYP52A3B 2409 CCCACTTAGATCTGTCTATTATGGGCTGTATGCTGTCTGATTCAGACCAAGAAAGATGGTTTGAAGCCATC 2478
CYP52A3A 1716 CTCAGACAAATCTGTCTATTATGGGCTGTATGCTGTCTGATTCAGACCAAGAAAGATGGTTTGAAGCCATC 1785
CYP52A3B 2269 CTCAGACAAATCTGTCTATTATGGGCTGTATGCTGTCTGATTCAGACCAAGAAAGATGGTTTGAAGCCATC 2338
CYP52D4A 1980 CACACTTGAAATGAGAAAGGTATATGGGAATGATAGCCATGTGTTTCGACCGGAGAGATGGGCTGGCTTAGA 2049
* * * * *

CYP52A1A 2535 TACTAAGAAAGTTGGGCTGGGCTTATGTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2604
CYP52A2A 2551 GACAAAGAAAGCTTGGGCTGGGCTTCTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2620
CYP52A2B 2424 GACAAAGAAAGCTTGGGCTGGGCTTCTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2493
CYP52A3A 2532 AACTAGAAAGTTGGGCTGGGCTATGTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2601
CYP52A3B 2319 AACTAGAAAGTTGGGCTGGGCTATGTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2388
CYP52A3A 2440 AACCAAAAGCTGGGCTGGGCTTACTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2509
CYP52A3B 2479 AACCAAAAGCTGGGCTGGGCTTACTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2548
CYP52A3A 1756 AACTAGAAAGCTTGGGCTGGGCTTCTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 1825
CYP52A3B 2339 AACTAGAAAGCTTGGGCTGGGCTTACTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2408
CYP52D4A 2050 GGGCAAGAGTTTGGGCTGGTGTATGTTCCATTCAAGCGTGGTCCAGAGTCTGCTTGGGTCAACAAATC 2119
* * * * *

CYP52A1A 2605 GCCTTGACTGAAGCTTCTTATGTGATCACTAGATTGGCCCAAGTGTGTTGAAGTGTCTCATCTGATCCAG 2674
CYP52A2A 2621 GCCTTGACAGAAAGCTTCTTATGTGATCACTAGATTGGCCCAAGTGTGTTGAAGTGTCTCATCTGATCCAG 2690
CYP52A2B 2494 GCCTTGACAGAAAGCTTCTTATGTGATCACTAGATTGGCCCAAGTGTGTTGAAGTGTCTCATCTGATCCAG 2563
CYP52A3A 2502 GCCTTGACAGAAAGCTTCTTATGTGATCACTAGATTGGCCCAAGTGTGTTGAAGTGTCTCATCTGATCCAG 2571
CYP52A3B 2389 GCCTTGACTGAAGCTTCTTATGTGATCACTAGATTGGCCCAAGTGTGTTGAAGTGTCTCATCTGATCCAG 2453
CYP52A3A 2510 GCCTTGACAGAAAGCTGGGCTATGTTGTTGATGATTGGTCCAGAGTGTCTCCACAGTTAGGGCTGGACCCAG 2579
CYP52A3B 2549 GCCTTGACAGAAAGCTGGGTTACGTTTGGTTCAGATTGGTCCAGAGTGTCTCCACAGTTAGGGCTGGACCCAG 2618
CYP52A3A 1856 GCTTTGACTGAAGCCCGGTTACGTTTGGTTAGACTTGTTCAGAGTGTTCACAACTTGTTCACAGACCCCG 1925
CYP52A3B 2409 GCTTTGACAGAAAGCCCGTTACGTTTGGTTAGACTTGTTCAGAGTGTTCACAACTTGTTCACAGACCCCG 2478
CYP52D4A 2120 CCAATCCTTGAAGCTTCTTATGTGATCACTAGATTGGCCCAAGTGTGTTGAAGTGTCTCATCTGATCCAG 2186
* * * * *

```

Figure 15H

CYP52A1A 2675 GTCTCGAATACCCCTCCACCAAAAGTGTATTCACTTCACCATGAGTCACAAAGATGGTGTCTTTGTCAAGAT 2744
CYP52A2A 2691 ACACCGAATATCCACCTAAGAAATGTTCGCAATTGACCATGTCCCTTTTCGACGGTGCCAAATATTGTAT 2760
CYP52A2B 2664 ACACCGAATATCCACCTAGGAAATGTTCGCAATTGACCATGTCCCTTTTCGACGGTGCCAAATATTGTAT 2633
CYP52A3A 2672 ACACCGAATATCCACCAAAATTCAGAAACACCTTGACCTTGTCCCTCTTTGATGGTGTCTGATGTTAAGAT 2741
CYP52A3B 2459 ACGCTGAGTACCCACCAAAAGGTTGACCAACTTGACCATGTGTTTGACGGATGGTGTATTGTCAAGTT 2649
CYP52A5A 2580 ACGAGGTGTACCCGGCCAAAGAGGTTGACCAACTTGACCATGTGTTTGACGGATGGTGTATTGTCAAGTT 2688
CYP52A5B 2619 ATGAAGTGTATCCACCAAAAGGTTGACCAACTTGACCATGTGTTTGACGGATGGTGTATTGTCAAGTT 2688
CYP52A8A 1926 AAACCAAGTACCCACCACTAGATTGGCACTTGACCATGTGCTTGTGTTGACGGGGATACGTCAGAT 2548
CYP52A8B 2479 AAACCTGAGTACCCACCACTAGATTGGCACTTGACCATGTGCTTGTGTTGACGGGGATACGTCAGAT 2548
CYP52D4A 2187 CTACCGAGTACCCACCAAAAGTGTTCATCTCAGCATGAGTCTTCTCAACGGGGTGTATCTCCGAA 2256
* * * *

CYP52A1A 2745 GTAA-AGTAGTCGATGCTGGGTATTGATTACATGT--GTATAGGAAGATTTGGTTTTTTATTGTTCT 2811
CYP52A2A 2761 GTATTAGAGGGTCATGTGTTATTTT-GATTGTTTA----GTTTGTAACTACTGATTAGGTTAAATTCATG 2824
CYP52A2B 2634 GTATTAGAGGATCATGTGTTATTTTGGATTGGTTAGTCTGTTTGTAGCTATTGATTAGGTTAAATTCACG 2703
CYP52A3A 2742 GTACTAAGGTTGCTTTTCTTGTCTAATTTTCTTGTATAGCTTGTGTTTAAATTTGAATCGGCAATTG 2811
CYP52A3B 2529 GTTCTAAGGTTGCTTTATCCTTGCTAGTGTATT---TATAGTTTGTATTTAAATTTGAATCGGCAATTG 2595
CYP52A5A 2650 TGACTACGGCGGTGGTGAATGCGTTTGATTGTTGA---GTTTCTGTTTGACGTAATGAGATAACTATTCA 2716
CYP52A5B 2689 TGACTAGTA-CGTA-TGAGTGCCTTTGATTGTTGA---GTTTCTGTTTGACGTAATGAGATAACTATTCA 2753
CYP52A8A 1996 GTCATAGGTTTCCC---CATACAAGTAGTTTACGTA---ATTATACACTGTTTCTCTTCATACC 2059
CYP52A8B 2549 GCAATAGGTTT-----TGTTTGACTTTGTTTCCATA-- 2580
CYP52D4A 2257 TAGAAGTGAATTATGTGTTTATGGTTAAATCGGGGCAAGCACTGCAAGTCATTGATGTTTGTGGAAGCCC 2326

CYP52A1A 2812 TTTTTTAAATTTTGTAAATTAG-TTAGAGATTTCATTAAATACATAGATGGGTGCTATTTCGGAAGT 2880
CYP52A2A 2825 GATTGTATTATTATGATAGGGGTT-----TGCGGTGTTGCAATCACTTGGGATCGTTCCAGGTTG 2885
CYP52A2B 2704 GATTGTATTATTATGATAGGGGGTGCCTGTGTGTGTGTGTGTGCAATTCACATGGGATCGTTCCAGGTTG 2773
CYP52A3A 2812 ATTTTCTGATACCAATAACCGTA-----GTGCAATTGACCAAAACCGTTCAAACTTTTGTCTC 2873
CYP52A3B 2596 ATTTTCTGTTACTAATAACTGTA-----GTGGGTTTGACCAAAACCGTTCAAACTTTTGTCTC 2657
CYP52A5A 2717 GATAAGGCGAGTGGATGTACGTTT-TGTAAAGATT---CCT-TACAACTTGGTGGG-TGTGTAGGTT 2781
CYP52A5B 2754 GATAAGGCGGGTGGATGTACGTTT-TGTAAAGATT---CCT-TACAACTTGGTGGG-TGTGTAGGTT 2817
CYP52A8A 2060 AAATGGACAAAAGTTTAAAGCATG-CCTAACCACTGACCG-GACAAATGTGTGCACTAGATTGTAAACA 2127
CYP52A8B 2581 -----TGCAAGT 2587
CYP52D4A 2327 AGCATTTGGTGTTCGGGAGCATCAATAACCAATGTCTGAAAGGTTTGTATTTCTTGACCTTCTTCTCT 2396

CYP52A1A 2881 TTACTTCTATCC--CCTGTATCCCTTATTATCCCTCTCAGTCACATGATTGCTGTAATTGTCGTGCAAGG 2948
CYP52A2A 2886 ATGTTTCCCTTCCATCCT--GTGAGTCAAAAGGAGTTTGTGTTTGTAACTCCGGACGATGTTTAAATAG 2953
CYP52A2B 2774 TTGTTTCCCTTCCATCCT--GTGAGTCAAAAGGAGTTTGTGTTTGTAACTCCGGACGATGTTTAAATAG 2841
CYP52A3A 2874 TCGTTGACG-----TGCTCGCTCATCAACACTGTTTGAAGACGAAAGA-GAAAATTTTGTGTA 2930
CYP52A3B 2658 TTTTCTTCCCTTACCTTCTGTTGCTCGCTCATCAGCACTGTTTGAAGACGAAAGA-GAAAATTTTGTGTA 2727
CYP52A5A 2782 GAGGTGCACTCTT-GGGAGATTACACCTTTTG-CAGCTCTCGGTATACACTTGTACTCTTGTAACTC 2849
CYP52A5B 2818 G----CATCTTAG-GGAGAGATAGCACCTTTTG-CAGCTCTCGGTATACACTTGTACTCTTGTAACTC 2881
CYP52A8A 2128 ATTGTAATAATAG-TGTACACTAATTGTGGTGGCGGAGATAAATTACAGTTTGGTTTGTGTAACTC 2196
CYP52A8B 2588 AGTTCAGTAAT--TACACACTAATTGTGGTGGCGGAGATAAATTACAGTTTGGTTTGTGTAAAT 2554
CYP52D4A 2397 GAGCTTCTTCCG--TCAAACTTGTACAGAAATGGCCATCTTCAGGAACAACCA-CGTACGACGGCGG 2463

CYP52A1A 2949 CACAACTCCCTAACGGACTTAAACCATAAACAGCTCAGAACCATAAAGCCGACATCACTCTCTTCTC 3018
CYP52A2A 2954 AAGGTGATCTCCATGTGATTGTTTGAAGTGTACTGTGATTATGTAATCTGCG-----GACGTTATA 3016
CYP52A2B 2842 AAGGTGATCTCCATGTGATTGTTT-GACTGCTACTCTGATTATGTAATCTGTAAGCCTAGACGTTATG 2910
CYP52A3A 2931 AACAACTGTGTCAAATTTACCAACGTTGAACCATATG--CAATGAGCGGCC-----CTTCAA 2989
CYP52A3B 2728 AACAACTGTGTCAAATTTACCAACGTTGAACCATATG--CAATGAGCGGCC-----CTTCAA 2788
CYP52A5A 2850 TATCAATCATGTGGGGGGGGGTTTCAATGTTTGGC-CATGTTGTTGATGTTAAATCCGCC-ACTACC 2917
CYP52A5B 2882 TGCCAAATCATGTGG-----GGATTCAATGTTTGGC-CATGTTGTTGATGTTAAATCCGCC-ACTACC 2944
CYP52A8A 2197 GCGGATATCTCTGGC-----AGTTTCTCTCTCCGC-AGCACTTTGCCACGGGTTTGTCTGCGGCCAA 2260
CYP52A8B 2655 TCGGACATCTCTGGT-----GGTTTCCCTTCTCCGC-AGCACTTTGCCACGGGTTTGTCTGCGGCCAA 2718
CYP52D4A 2464 TACCGCATCTGGAATA---TCTCGCGCTGTTCAAGTAG--CACGAAACAGCAACGACGTCAACATCTG 2528

Figure 15I

CYP52A1A 3019 TCTTCTCCAACCAATAGCATGGACAGACCCACCCCTCTATCCGAATCGAAGACCCCTTATGACTCCATAC 3088
 CYP52A2A 3017 CAAGCATGTGATTTGTGGTTTT-----GCAGCCT-TTTCACGACAAATGATCGTACAGCGATTACGTAA 3079
 CYP52A2B 2911 CAAGCATGTGATTTGTGGTTTT-----GCAGCCT-TTTCACGACAAATGATCGTACAGCGATTACGTAA 2975
 CYP52A3A 2990 CTGGTCGCTGGGAAGCATTCCGG-----GATATCTACAACGCCCTTAAGTTTGAGGAAGACATTGATTAG 3054
 CYP52A3B 2789 CTGGTCACCTGGAGGCATTCCGG-----GATATCTACAACGCCCTTAAGTTTGAGGAAGACATTGATTAG 2853
 CYP52A5A 2918 CAATCTCAGATGAAGCTCAAGCACACTAAAAAAGATGTTGGGGGAAACTT-TGGTTTCCTTC 2986
 CYP52A5B 2945 CAATCTCAGATGAAGCTCAAGCACACTAAAAAAGATGTTGGGGGAAACTT-TGGTTTCCTTC 3005
 CYP52A8A 2261 CAATTCAGAAAGGGGG-----AGAACTTAACACCCCTTATCTCTCCACTC-TAGGTTGTAGCT 2318
 CYP52A8B 2719 CAATTCAGAAAGGGGGGGGGGGGGGAGAAAGTTAACACCCCTTGTTC--CACCG-TAGGTTGTAGCT 2785
 CYP52D4A 2529 CTTCCCAATCTTGAACCC-----ACAGATACCCCTGCGGCTTCATGGATCAAAAACGTGGCAACC 2590

CYP52A1A 3089 CCACCTGGAGCCCTCAAGCCACACAGTCATCCAGCCACCCATCACCACATCCCTCTACTCGACAAC 3158
 CYP52A2A 3080 TCTTTGTTA-----GAGGGGTAACAAAAAACAATGCGAGCCAGAAATTCAGATTCGCAACAAATG 3144
 CYP52A2B 2976 TCCATATTAT-TTGAAGGGGTAATAAAAAATAA-TGGCAGCCAGAAATTCAGATTCGCAACAAATG 3043
 CYP52A3A 3055 ACACCATAGA-TTTCAGCGGCATCAAGAAATGACC-----TTGCCACATTTTGAACACCCCAACCCACTG 3119
 CYP52A3B 2854 ACACCATAGA-TTTCAGCGGCATCAAGAAATGACC-----TTGCCACATTTTGAACACCCCAACCCACTG 2918
 CYP52A3B 2987 TTAGTAAT-AAAACCTCTCACTCTCACTCTCACTCTCACTCTCACTCAGACAAACCAACCCCTGGGCTGC 3054
 CYP52A5A 2987 TTAGTAAT-AAAACCTCTCACTCTCACTCTCACTCTCACTCTCACTCAGACAAACCAACCCCTGGGCTGC 3073
 CYP52A5B 3006 TTGGGGAAA--ACTTTCGTTTCTTCTCACTAAATAAGCTTCTCACTCAGACAAACCAACCCCTGGGCTGC 2386
 CYP52A8A 2319 CTGTGGGG--ATGCAATGTGCTACGTTTTTATGTTTTGTCTAGACTTTGATGATTAGTTGGATTTC 2854
 CYP52A8B 2786 CTGTGGGGGATGTAAATGTGCTACGTTTTTC-ATGTTTGGCCCACTTTGATGATTAGTTGGATTTC 2658
 CYP52D4A 2591 CCGCGTATATGTCATGTAAATCTCCATGGCCACT--CCATCAACACACTGATGGAGCGACTGACGGTG 2658

CYP52A1A 3159 GTCCAAAGACGGCGAGTTCTGGTGTGCCCGGAAATCAGCCATCCCGGCCACATACAAAGCAGCGTTGATT 3228
 CYP52A2A 3145 CAAAAAATGGGAAACTC--CAACAGACAAA-AAAAAATCCGACGACTCCGAACCCACAGAACAAATG 3211
 CYP52A2B 3044 CAAAAGATGAGAACTC--CAACAGAAAAATAAAAAATCCGACGACTCCGAACCCACAAACAAATG 3111
 CYP52A3A 3120 GAAGAAATCAGCCAGA-----AACTAGGCGATGGATCCAGGCTGTGACCTTGCCCAATGGAGACGAAGTG 3185
 CYP52A3B 2919 GAAGAAATCAGCCAGA-----AACTAGGCGATGGATCCAGGCTGTGACCTTGCCCAATGGAGACGAAGTG 2984
 CYP52A5A 3055 AGACAAACCAAGAAAAAAGAAACAAATCCAGATAGAAAAAAGAGGGCT-GGACAAACCAATAAAT-AAAC 3122
 CYP52A5B 3074 AGACAAACCAAGAAAAA-----CAAAATCCAGATAGAAAGAAAGAGGGCT-GGACAAACCAATAAAT-AAAC 3135
 CYP52A8A 2387 TTATGTCTGAGGCGTG-----CTTGACACAAAGTGCAAAAGGTGACAGGCG-ACGCTATTGACAT-GAAC 2450
 CYP52A8B 2855 TTATGTCTAAGGCGTG-----CTTGACACAAAGTGCAAAAGGTGACAGGCG-ACGCTATTGACAT-GAAC 2918
 CYP52D4A 2659 CACCACTGCCCTCGG-----TTGAGTCAAGGCGATGATGCCGGATCCAGTACTCCAAATGGAACC 2722

CYP52A1A 3229 CGGTGCATACTCGGCGAGCCACAAATGTAAGCCAGCATTCCGACCATGAAGCAAGTACATTACAGGA 3298
 CYP52A2A 3212 GGG-----CGCCAGAAATTTGACTATTGTGACTTTTTTA-----CGCTAACGCTCAATGCAATG 3266
 CYP52A2B 3112 GGG--GCGCCAGAAATTTGACTATTGTGACTTTTTTT-ATTTTTCCGTTAACTTTCAATGCAATG 3177
 CYP52A3A 3186 GAGTTGAACCAAGCGTTCTAGAAAGTTACCACTATTGTGCAAGTGAATTTGACTTGGACCAATGAACG 3255
 CYP52A3B 2985 GAGTTGAACCAAGCGTTCTAGAAAGTTACCACTATTGTGCAAGTGAATTTGACTTGGACCAATGAACG 3054
 CYP52A5A 3123 AATCTAGGCTCTACTCCATCTTCACTGTTCTTCTTCTCAGACTTAGCT-AAACAAACCTCACTTCA 3191
 CYP52A5B 3136 AACCTAGGCTCCACTCCATCTTCACT---TCTTCTTCTCAGACTTAGCT-AAACAAACCTCACTTCA 3201
 CYP52A8A 2451 GCGAAAGGGTATTGTCATCAATACGAG--GGGCTGACTCTAGTCTAGG---ATGGCAGTCTAGGTTGC 2525
 CYP52A8B 2919 GCAAAAGGGTAAATTTGATCGATACGAG--GGGTTGCCCTCTGGTCTAAG---AAGGACCCCAAGTTGC 2983
 CYP52D4A 2723 TCT-----GCACGGTGTGCTGCACTTTTGAAGCGTATTTCGA-----TCCATGATCGTCTTTGG 2779

CYP52A1A 3299 TCACGGGTGTTTCAG-TGTCCGAGATTGAGAAAGTTGACGATGGATGGAAGTACGATCTCGTTGCGGATT 3367
 CYP52A2A 3267 TAGTGCGTCTTACAGG-----GGTATTGCTTCTCAATGCAAGGCA-CAGTTGAAGGTTTGCAACC 3328
 CYP52A2B 3178 AAGTGTTTACAGGGGTGGTATGGTGGTTTCTCAATGCAAGGCA-CAGTTGAAGGTTTGCAACC 3246
 CYP52A3A 3256 CGGCGAGAGTTGTTATACTA-CGCTGGCGACATATCTCAAGAAAGGCAATCAATCGCAGACAGTGCCA 3324
 CYP52A3B 3055 CGGCGAGAGTTGTTATACTA-CGCGCGGACATATCTCAAGAAAGGCAATCAATCGCAGACAGTGCCA 3123
 CYP52A5A 3192 CCATGGATTACGAGGCATCAGCGGTGCTCCATCAGAGG-CGAGGCTTGAAGAACTCG--CAGAAAT 3258
 CYP52A5B 3202 CCATGGATTACGAGGCATCAGCGGTGCTCCATCAGAGG-CGAGGCTTGAAGAACTCG--CAGAAAT 3268
 CYP52A8A 2516 AAACATGTTGACCA-TATCCCTCTGAGTTGGTGCAC--CTGCGCTAGCC-ACCCTCA--GCGATCG 2579
 CYP52A8B 2984 AAACATGTTGACCTG-CATCCCTCTGAGTTGGTGCAC--CAGGCTACGCTTACCTCA--GCGATCG 3048
 CYP52D4A 2780 TGCTGTAGTATAACGAGCT--CTTGCTGCTTGAATGCAAGGTTGATGTTGTTGATGTTGCT 2847

Figure 15J

CYP52A1A 3368 ACGACTTCGGTGGGTTGTTATCTAAACGAAGATTCTATGAGACGCAGCATGTGTTTCGGTTCGAGGATTG 3437
CYP52A2A 3329 TAACGTTGCCCCGGTGTCAACTCAATTGAC-----G--AGTAACTTCCTAAGCTCGAATTATGC 3385
CYP52A2B 3247 TAACGTTGCACCATATCAACTCAATTATC-----CTCATTCTATGTGATAAAGAACAGCCAAA 3305
CYP52A3A 3325 GATTGCTTTATTATTGAGAGCAAACTAC-----ATCTTGAACATACTTGGGTACTTTAT 3379
CYP52A3B 3124 GATTGCTTTACTATTGAGAGCAAACTAC-----ATCTTGAACATACTTGGGTACTTTAT 3178
CYP52A5A 3259 G-ACCATCCAGAACCAGCCATCCAGCT-----TGAAAGAAATCAACACCGGCATCCAGAAGGACGACTT 3321
CYP52A5B 3269 G-ACCATCCAGAACCAGCCATCCAGCT-----TGAAAGAAATCAACACCGGCATCCAGAAGGACGACTT 3331
CYP52A8A 2580 GCACTTTCCGTTGTTCAATATTCTC-----CTTCCCATTTGTTCCAGGGGTTA--TC 2629
CYP52A8B 3049 GCACTTTCCGTTGCTCAATATTCTCT-----CCCCCTGCTTCCCCCATTTGTTCCAGGGATTA--TC 3110
CYP52D4A 2848 GCGTGCTTGGTTGCAAGTCTTCGATCG-----ACGCTAGTGAGTAGACAGTTGGCGGG 2901

CYP52A1A 3438 TGGCTACGTCATGAGTGTGCTTTTGTATGGACCCAAAGGAGGAAGGTTACGTGGTGGGACGTACAGATCC 3507
CYP52A2A 3386 AGCT-CGTGCGTCAACCTATGTGACGAAAGAAATCCAAAAA--AATCGAAA-ATGCGACTTTCGAT 3451
CYP52A2B 3306 AGGT-AAT-TGGCAGACCCCCAAGGGGAACACGGAGTAGAAAGC--AATGGAACACGCCCATGACAGT 3371
CYP52A3A 3380 TTCG-AAGCAGCGATTGGATTGTATGATGACGGACAAACGACGCGT--TGTTTGATAGTATTTTGAAAGT 3446
CYP52A3B 3179 TTCG-AAGCAGCGATTGGATTGTATGATGACGGACAAACGACGCGT--TGTTTGATAATATTTTGAAAGT 3245
CYP52A5A 3322 TGCC-AAGTGTGTTGCTGCCACCCCGAAATCCCAACCAAGCACA--AGTTGAACGGCAACACGAATT- 3387
CYP52A5B 3332 TGCC-AAGTGTGTTGCTGCCACCCCGAAATCCCAACCAAGCACA--AGTTGAACGGCAACACGAATT- 3397
CYP52A8A 2630 AACA-ACGTTGCCGGCTCCTC-----CCCAATTA-----CAAGAAAATTAAT- 2674
CYP52A8B 3111 AACA-ACGTTGCCGGCTCCTC-----CCCAATTA-----GTACAAGAAAATTAAT- 3171
CYP52D4A 2902 GGTGGTGCTCGGGCTTATCTGTGTTGTGTTCTCTTAGT--CTTGAATGACGCTGTTATCGAC 2969

CYP52A1A 3508 ATTGAAGGTTGAGCTGGGGTAAAGACGGGACGTGGA-GTGGACCATGG---CGACGACGTGGATCCT 3573
CYP52A2A 3452 TTTGAATAAACCAAAAGAAATGTGCGACTTTTTTC-----TCGCTCTCGCTCTCTCGACCCAAATCA 3516
CYP52A2B 3372 GCCATTAGCCCAACA--ACACATCTGATTTCTTTT-----TTTTTTGTGCGCAGGTGACACCTGG 3433
CYP52A3A 3447 TTTGAAAAGATCTAC-----AAGTTGATAAGCGGTGTTGA-----ACGATATGATTGACAAGCAAAAGGTGA 3507
CYP52A3B 3246 TTTGAAAAGATCTAC-----AAGTTGATAAGCGGTGTTGA-----ACGATATGATTGACAAGCAAAAGGTGA 3306
CYP52A5A 3388 GTCTGAGGTGCGCATTTGCCAAAAGGAGTACGAGGTGTTGATTGCTTTGAGGCGACGCCACAAAAGACCCA 3457
CYP52A5B 3398 GTCCGAAGTGCCTTATGCCAAAAGGAGTACGAGGTGTTGATTGCTTTGAGGCGACGCCACAAAAGACCCA 3467
CYP52A8A 2676 GTCCGACGGCAGCATCTGTCAAAGATACAGATAA-----ACCTTAAATCTGCAAAAACAAGACCCC 2736
CYP52A8B 3172 GTCCGACGGCAGCATCTGTCAAAGATACAGATAA-----ACCTTAA-----TCC 3216
CYP52D4A 2970 GGTTCGTAGTATAAGTAGCGCAATATGAGATGTATA-----TCCGATCACCAAGACTCTTCAGCCT 3034

CYP52A1A 3574 GGTGGGTTTATCCCGCA-ATGGATAACTCGATTGAGCA-TCCCTGGAGCAATCGCAAAAGATGTGCTAG 3641
CYP52A2A 3517 CAACAAATCCTCGCGCGCAGTATTTGACGAAAC--CACAAACAAATAAAAAACAAATCTACACCACT 3584
CYP52A2B 3434 ACTTTAGTTATTGCCC-CATTAAGTTAACAATCT--CACCTTTGGCTCTCCCACTGTCTCCGCTCCA 3500
CYP52A3A 3508 CAAGCGACATCAACAGTCTAGCATTCAATTTG--CATCACTACTCGAGAGGTCAACTATTCTCCGCA 3575
CYP52A3B 3307 CAAGCGACATCAACAGTCTAGCATTCAATTTG--CATCACTACTCGAGAGGTCAACTATTCTCCGCA 3374
CYP52A5A 3458 ATCAAAGTGACCTCCAGATCAAGATCTTGATTGACAAAGTTCAAGGTGTACTTGT---TTGAGTTGCCCG 3524
CYP52A5B 3468 ATCAAAGTGACCTCCAGATCAAGATCTTGATTGACAAAGTTCAAGGTGTACTTGT---TTGAGTTGCCCG 3534
CYP52A8A 2737 TCCCATAGCCTAGAGGACAGCAAGATGATGGAGCAACTCTCCAGTACTGGTACATCCGACTCTCTG 2806
CYP52A8B 3217 CTCCATAGCCTAGAGGACATCAAAAGATGATTGAGCAACTCTCCAGTACTGGTACATTGCACTCCCTG 3286
CYP52D4A 3035 GTTACACGACTGAGGCTGTGTGGCGTGTGACCAATTGGTTCTTTGGTGACCTAGATTGGTCCGCGAGG 3104

CYP52A1A 3642 TG---TATTAACATACATACAGAAATAAAACGTTGCTTTGATTCAATTGGTTT---GGTCTTGTGGGTT 3705
CYP52A2A 3595 T---CTTTTCTTCCAGTCAACAAAAACAACAAATATACACCAATTTCACGATTTTTGGCTCTAT 3650
CYP52A2B 3501 TG---CTCGTTTACACCTTCGAGCTAACGACAAACACCCATGAGGGGAATGGGCAAGTT----- 3562
CYP52A3A 3576 CA---CGAATTTTGGG-ACTGGTTTGTGTTGGATTGGTGCACATCTATTTCAACCAAGTTTGGGCAATTA 3641
CYP52A3B 3375 CA---CGAATTTTGGG-ACTGGTTTGTGTTGGATTGGTGCACATCTATTTCAACCAAGTTTGGGCAATTA 3440
CYP52A5A 3525 AC---CAGAAGTTCTCTACTCCATCGTGTCCAACTCCGTCACATCGCCCCC-TGGACCTTGCTCGGG 3590
CYP52A5B 3535 AC---CAGAAGTTCTCTACTCCATCGTGTCCAACTCCGTTAACATTGCCCCC-TGGACCTTGCTCGGG 3600
CYP52A8A 2807 TA---TGGTTCACTCTCGCTACTTGGCTTCCACGCAAGGCGCTCTACTTG-CGCCACAGCTCGGG 2872
CYP52A8B 3287 TA---TGGTTCACTCTCGCTACTTGGCTTCCACGCAAGGCGCTCTACTTG-CGCCACAGCTCGGG 3352
CYP52D4A 3105 GAAAGCAAGGGCTGCTAGGGGGCATACCAACAAAGTGTGTAATCAGTATCTATGGTGCTACCATGTG 3174

Figure 15K

CYP52A1A 3706 CCGAGCCAAATATTTCACATCATCTCCTAAATTTCTCCAAGAAATCCCAACGTAGCGTAGTCCAGCAGGCCCT 3775
 CYP52A2A 3651 AAATGCTATATAATGGTTTAAATCAACTCAGGTATGTTTAT-TTACTGTTTTCAGCTCAAGTATGT--T 3717
 CYP52A2B 3563 AAACACTTTTGGTTTCAATGATTCTATTGCTACTCTCTTGTGTTTGTGTTTGTGATTTCACCATGT--G 3630
 CYP52A3A 3642 GACAACTACAAGAAAGTATTGGCATTGATACTGAAGAACATCAGTGATGAAGATATCTTGATCGTAC--G 3709
 CYP52A3B 3441 GACAACTACAAGAAAGTATTGGCATTGATACTGAAGAACATCAGTGATGAAGATATCTTGATCGTAC--G 3508
 CYP52A5A 3591 AGAAGTTGACCACGGGCTTGATCAACTTGGGCTTCCAGAAACAACAGCAGCACTTGGACGAGGTCAAT--G 3659
 CYP52A5B 3601 AGAAGTTGACCACGGGCTTGATCAACTTGGGCTTCCAGAAACAACAGCAGCACTTGGACGAGGTCAAT--G 3659
 CYP52A8A 2873 CGGCGCCATTACGCAACCCAGTACGACGGCTGGTATGGGTTCAAGTTTGGGCGGGAGTTTCTCAA--G 2940
 CYP52A8B 3353 CGGCGCGCTTACGCAACCCAGTACGACGGCTGGTATGGGTTCAAGTTTGGGCGGGAGTTTCTCAA--G 3420
 CYP52D4A 3175 TGTGGTTGGGGGAAATTCGCCCATTTTGTGTAAAGAAAGTTCTAGAAAGTTCTCGTGGGTTCTGAG--A 3243

 CYP52A1A 3776 CTGAGATCTTATTAAATATCGACTTCTCAACCACGGTGGAAATC--CCGTTACAGACCATTTGTTACCTGTA 3843
 CYP52A2A 3718 CAAATACTAACTACTTTTGATGTTTGTGCGCTTTCTAGAATCAAAACAACGCCCAACGCGGAGCTT 3787
 CYP52A2B 3631 AAATAAACGACAATTATATATACCTTT---TCGTCTGTCTC---CAATGTCT-CTTTTGTGCTGCCATT 3692
 CYP52A3A 3710 CTTCCTCCCATCGACACTACAATTGTTTAAAGCTGGTGTGGACAA-GAAAGACGACGCTGCAAGTGAACA 3778
 CYP52A3B 3509 CTTCCTCCCATCGACACTACAATTGTTTAAAGCTGGTGTGGACAA-GAAAGACGACGCTGCAAGTGAACA 3577
 CYP52A5A 3660 ACATCTTCAACGAGTTTATCGACAAGTTCTTTGGCAACACAGAG--CCGCAATTGAC-----CAACTTCT 3722
 CYP52A5B 3670 ACATCTTCAACGAGTTTATCGACAAGTTCTTTGGCAACACAGAG--CCGCAATTGAC-----CAACTTCT 3732
 CYP52A8A 2941 GCGAAGAAAGATCGGGCGGCAACGGACTTGGTGATGCGCGGTT--CCGTGGCGG-----CATGGACA 3001
 CYP52A8B 3421 GCGAAGAAAGATCGGGCGGCAACGGACTTGGTGATGCGCGGTT--CCGTGGAGGGG-----CATGGATA 3484
 CYP52D4A 3244 ATCTGCTG3AACCATCCACCCGCAATTCGGTTGCCAAAGTGGGAA-GAGCAATCAACCCACCTGCTTGG 3312

 CYP52A1A 3844 GTGTGTTTGTCTTGTCTTGATGACAAATGATGATTGTCACGATACCTGAATAATAAATCATCCAGT 3913
 CYP52A2A 3788 GTCGAATAGACGGTTTGTGTTACTCATTAGATGGTCCAGATTACTTTTCAAGCCAAAGTCTCT-CGAGTT 3856
 CYP52A2B 3693 TGTCTTTTGTCTTTTGTCTTTGCACT---CTCTCCCACTCCCAATCAGTGCAAGCAACA-CAA 3755
 CYP52A3A 3779 GTTCTACAAGTACATCACTCAACAGT--GTGCAAGACTACAACCTCAACATCGGCTCCACAGCCAAAG 3846
 CYP52A3B 3578 GTTCTACAAGTACATCACTCAACAGT--GTGCAAGACTACAACCTCAACATCGGAGCCACAGCCAAAG 3645
 CYP52A5A 3723 TGACCTTGTGCGGTGTGTTGGACGGGTGATTGACCAATGCC-AACTTCTTGAAGCGTGTCTCCAGGACCT 3791
 CYP52A5B 3733 TGACCTTGTGCGGTGTGTTGGACGGGTGATTGACCAATGCC-AACTTCTTGAAGCGTGTCTCCAGGACCT 3801
 CYP52A8A 3002 CTTTCTCGAGCTACACTTTTCGGCATCCATATCATCTTACC-CGGGACCCGGAGAACATCAAGGCGGTCT 3070
 CYP52A8B 3485 CTTTCTCGAGCTATACTTTTCGGCATCCATATCATCTTACC-CGGGACCCGGAGAACATCAAGGCGGTCT 3553
 CYP52D4A 3313 CCCAATCAGCCATTCCCTGCGAATATATAATTCAGC 3348

 CYP52A1A 3914 CATTGAGCTTATTACTCGTGAACCTATGAAGAACTCATTCAAGCCGTTCCCAAAAAACCGAATTGAA 3983
 CYP52A2A 3857 TTGTTTGTCTGTTTCCCAATTCCTAATATGAAGGTTTTATAGGTCCAAAGACCCCAAGGCATAGTT 3926
 CYP52A2B 3756 3755
 CYP52A3A 3847 ATGATATCGATTGTCCTCAAAACCAAACTCAGTGGCTTTGAGGTGTTGACGAGTT 3900
 CYP52A3B 3646 ATGATATCGATTGTCCTCAAAAGCC 3663
 CYP52A5A 3752 TCAAGATCTTCTTGAACCTTGGACTCGTATGTGGAC 3826
 CYP52A5B 3802 TCAAGATCTTCTTGAACCTTGGACTCGTATGTGGAC 3871
 CYP52A8A 3071 TGGCGACGCAATTCGATGACTTCTCGCTGGTGCAAGATCAGGTTCTTGAAGCCGTTGTTGGGGTATGG 3140
 CYP52A8B 3554 TGGCGACGCAATTCGATGACTTCTCGCTGGTGCAAGATCAGGTTCTTGAAGCCGTTGTTGGGGTATGG 3579
 CYP52D4A 3149 3148

 CYP52A1A 3984 GATCTTGCTCAACTGGTCAAGCAAGTAGATCGCCATGATCTGATACTTTACCAAGCTATCTCTCCA 4053
 CYP52A2A 3927 TTTTGGTTCCCTTCTTGTGCTG 3948
 CYP52A2B 3756 3755
 CYP52A3A 3901 3900
 CYP52A3B 3669 3668
 CYP52A5A 3827 3826
 CYP52A5B 3872 CGACTTTTGTACGACGAGCCGAACGAGTACCAGAACTT 3910
 CYP52A8A 3141 GATATTACGTT 3152
 CYP52A8B 3580 3579
 CYP52D4A 3349 3348

Figure 15L

CYP52A1A	4054	AGTTCTCCACGTACGGCAAGTACGGCAACGAGCTCTGGAAGCTTTT	4115
CYP52A2A	3949		3949
CYP52A2B	3756		3756
CYP52A3A	3901		3900
CYP52A3B	3669		3668
CYP52A5A	3827		3826
CYP52A5B	3911		3910
CYP52A8A	3153		3152
CYP52A8B	3580		3579
CYP52D4A	3349		3348

Figure 15M

CYP52A1A	1	MATQEIIDSVLPLYL-----TKWYTVITA AVLVLFLISTNIKNYV	38
CYP52A2A	1	MTVHDIIATY-----FTK WYVIVPLALIA YRVLDYFYGRY	35
CYP52A2B	1	MTAQDIIATY-----ITK WYVIVPLALIA YRVLDYFYGRY	35
CYP52A3A	1	MSSSPSFAQEV LATTSPYIEYFLDN YTRWYFIPLVLLSLNFISLLHTRY	50
CYP52A3B	1	MSSSPSFAQEV LATTSPYIEYFLDN YTRWYFIPLVLLSLNFISLLHTRY	50
CYP52A5A	1	MIEQLLEY-----WYV VVPVLYI IKQLLAYTKTRV	30
CYP52A5B	1	MIEQILEY-----WYIV VPVLYI IKQLIAYS KTRV	30
CYP52A8A	1	MLDQILHY-----WYIVL PLLAIINQIVAHVRTNY	30
CYP52A8B	1	MLDQIFHY-----WYIVL PLLV IIKQIVAHARTNY	30
CYP52D4A	1	MAISSLLSWD-----VICVVFICVCVYFGYECYTKY	32
CYP52A1A	39	KAKKLKCVDP PPKYK DAGLTGILSLIAAIKAKNDGR LANFAD---EVFDEY	85
CYP52A2A	36	LMYKLGAKPFFQKQTDGCFGFKAPLELLKKKSDGTLIDFTL---QRIHDL	82
CYP52A2B	36	LMYKLGAKPFFQKQTDG YFGFKAPLELLKKKSDGTLIDFTL---ERIQAL	82
CYP52A3A	51	LERRFHAKPLGNVVRDPTFGIATPLLLIYLKSKGTVMKFANGWLWNNKYIV	100
CYP52A3B	51	LERRFHAKPLGNVVRDPTFGIATPLLLIYLKSKGTVMKFAWSFWNNKYIV	100
CYP52A5A	31	LMKKLGAA PVTNKLYDNAFGIVNGWKALQFKKEGRAQEYND---YKFDHS	77
CYP52A5B	31	LMKKLGAA PVTNQLYDNVFGIVNGWKALQFKKEGRAQEYND---HKFDSS	77
CYP52A8A	31	LMKKLGAKPFTHVQRD GWLGFKFGREFLKAKSAGRLVDLII---SRFHDN	77
CYP52A8B	31	LMKKLGAKPFTHVQLD GWLGFKFGREFLKAKSAGRQVDLII---SRFHDN	77
CYP52D4A	33	LMHKHGAREIENVINDGFFGFRPLPLLLMRASNEGR LIEFSV---KRFESA	79
CYP52A1A	86	PN--HTFYLSVAGALKIVMTVDPENIKAVLATQFTDFSLGTRHAFAPLL	133
CYP52A2A	83	DRPDIPTFTFPVFSINLVNTEPENIKAILATQFND FSLGTRHSHFAPLL	132
CYP52A2B	83	NRPDIPTFTFPIFSINLISTEPENIKAILATQFND FSLGTRHSHFAPLL	132
CYP52A3A	101	RDPKYKTTGLRIVGLPLIETMDPENIKAVLATQFND FSLGTRHDFLYSLL	150
CYP52A3B	101	KDPKYKTTGLRIVGLPLIETIDPENIKAVLATQFND FSLGTRHDFLYSLL	150
CYP52A5A	78	KNPSVGTYVSILFGTRIVVT KD PENIKAILATQFGD FSLGKRHLEFKPLL	127
CYP52A5B	78	KNPSVGTYVSILFGTRIVVT KD PENIKAILATQFGD FSLGKRHLEFKPLL	127
CYP52A8A	78	ED----TFSSYAFGNHVVFTRDPENIKALLATQFGD FSLGSRVKFFKPLL	123
CYP52A8B	78	ED----TFSSYAFGNHVVFTRDPENIKALLATQFGD FSLGSRVKFFKPLL	123
CYP52D4A	80	PHQNKXTLVNRALSVPVILT KD PVNIKAMLSTQFDD FSLGLRLHQFAPLL	129
CYP52A1A	134	GDGIFTLDGEGWKHSRAML RPQFARDQIGHVKALEPHIQIMAKQIKLNQG	183
CYP52A2A	133	GDGIFTLDGAGWKHSRSM LRPQFAREQISHVKLLEPHVQVFFKHVRKAQG	182
CYP52A2B	133	GDGIFTLDGAGWKHSRSM LRPQFAREQISHVKLLEPHVQVFFKHVRKAQG	182
CYP52A3A	151	GDGIFTLDGAGWKHSRTML RPQFAREQVSHVKLLEPHVQVFFKHVRKHG	200
CYP52A3B	151	GDGIFTLDGAGWKHSRTML RPQFAREQVSHVKLLEPHVQVFFKHVRKHG	200
CYP52A5A	128	GDGIFTLDGEGWKHSRAML RPQFAREQVAHVTSLEPHFQLLKKHILKHKG	177
CYP52A5B	128	GDGIFTLDGEGWKHSRSM LRPQFAREQVAHVTSLEPHFQLLKKHILKHKG	177
CYP52A8A	124	YGIFTLDAEGWKHSRAML RPQFAREQVAHVTSLEPHFQLLKKHILKHKG	173
CYP52A8B	124	YGIFTLDAEGWKHSRAML RPQFAREQVAHVTSLEPHFQLLKKHILKHKG	173
CYP52D4A	130	GKGIFTLDGPEWKQSRSM LRPQFAKDRVSHILDLEPHFVLLRKXHDGNG	179
* ***** **.*.*****. * . ***** . *.. *			

Figure 16A

CYP52A1A	1	MATQEIIDSVLPLYL-----TKWYTVITA AVLVLFLISTN IKNYV	38
CYP52A2A	1	MTVHDIIATY-----FTKWYVIVPLALIA YRVLDYFYGRY	35
CYP52A2B	1	MTAQDIIATY-----ITKWYVIVPLALIA YRVLDYFYGRY	35
CYP52A3A	1	MSSSPSFAQEV LATTSPYIEYFLDNYTRWYYFIPLVLLSLNFISLLHTRY	50
CYP52A3B	1	MSSSPSFAQEV LATTSPYIEYFLDNYTRWYYFIPLVLLSLNFISLLHTRY	50
CYP52A5A	1	MIEQLLEY-----WYVVVPVLYIIKQLLAYTKTRV	30
CYP52A5B	1	MIEQILEY-----WYIVVPVLYIIKQLIAYS KTRV	30
CYP52A8A	1	MLDQILHY-----WYIVLPLLAIINQIVAHVRTNY	30
CYP52A8B	1	MLDQIFHY-----WYIVLPLLVIKQIVAHARTNY	30
CYP52D4A	1	MAISSLLSWD-----VICVVFICVCVYFGYEYCYTKY	32
CYP52A1A	39	KAKKLKCVDP PPKYKLDAGLTGILSLIAAIKAKNDGR LANFAD---EVFDEY	85
CYP52A2A	36	LMYKLGAKPFFQKQTDGCFGFKAPLELLKKKSDGTLIDFTL---QRIHDL	82
CYP52A2B	36	LMYKLGAKPFFQKQTDGYFGFKAPLELLKKKSDGTLIDFTL---ERIQAL	82
CYP52A3A	51	LERRFHAKPLGNFVRDPTFGIATPLLLIYLKSKGTVMKFWANGLWNNKYIV	100
CYP52A3B	51	LERRFHAKPLGNVVDPTFGIATPLILIYLLKSKGTVMKFAWSFWNNKYIV	100
CYP52A5A	31	LMKKLGAAAPVTNKLYDNAFGIVNGWKALQFKKEGRAQEYND---YKFDHS	77
CYP52A5B	31	LMKKLGAAAPITNQLYDNVFGIVNGWKALQFKKEGRAQEYND---HKFDSS	77
CYP52A8A	31	LMKKLGAKPFTHVQRDGLGFKFGREFLKAKSAGRLVDLII---SRFADN	77
CYP52A8B	31	LMKKLGAKPFTHVQLDGLGFKFGREFLKAKSAGRQVDLII---SRFDN	77
CYP52D4A	33	LMHKHGAREIENVINDGFFGFRLLPLLLMRASNEGR LIEFSV---KRFESA	79
CYP52A1A	86	PN--HTFYLSVAGALKXIVMTVDPENIKAVLATQFTDPSLGTRHAFAPLL	133
CYP52A2A	83	DRPDIPFTFTFPVSINLVNTLEPENIKAILATQFNDPSLGTRHSHFAPLL	132
CYP52A2B	83	NRPDIPTFTFPFISINLISTLEPENIKAILATQFNDPSLGTRHSHFAPLL	132
CYP52A3A	101	RDPKYKTTGLRIVGLPLIETMDPENIKAVLATQFNDPSLGTRHDFLYSLL	150
CYP52A3B	101	KDPKYKTTGLRIVGLPLIETIDPENIKAVLATQFNDPSLGTRHDFLYSLL	150
CYP52A5A	78	KNPSVGTYVSILFGTRIVVTKDPENIKAILATQFGDFSLGKXHALFKPPL	127
CYP52A5B	78	KNPSVGTYVSILFGTRIVVTKDPENIKAILATQFGDFSLGKXHALFKPPL	127
CYP52A8A	78	ED----TFSSYAFGNHVVFTRDPENIKALLATQFGDFSLGSRVKFFKPPL	123
CYP52A8B	78	ED----TFSSYAFGNHVVFTRDPENIKALLATQFGDFSLGSRVKFFKPPL	123
CYP52D4A	80	PHPQNKTLVNRALSVPVILT KDPVNIKAMLSTQFDDPSLGLRLHQFAPLL	129
CYP52A1A	134	GDGIFTLDGEGWKHSRAMLRPQFARDQIGHVKALEPHIQIMAKQIKLNQG	183
CYP52A2A	133	GDGIFTLDGAGWKHSRSMRLRPQFAREQISHVKLLEPHVQVFFKHVRKAQG	182
CYP52A2B	133	GDGIFTLDGAGWKHSRSMRLRPQFAREQISHVKLLEPHVQVFFKHVRKAQG	182
CYP52A3A	151	GDGIFTLDGAGWKHSRTMLRPQFAREQVSHVKLLEPHVQVFFKHVRKHG	200
CYP52A3B	151	GDGIFTLDGAGWKHSRTMLRPQFAREQVSHVKLLEPHVQVFFKHVRKHG	200
CYP52A5A	128	GDGIFTLDGEGWKHSRAMLRPQFAREQVAHVTSLEPHFQLLKXHLKHG	177
CYP52A5B	128	GDGIFTLDGEGWKHSRSMRLRPQFAREQVAHVTSLEPHFQLLKXHLKHG	177
CYP52A8A	124	GYGIFTLDAEGWKHSRAMLRPQFAREQVAHVTSLEPHFQLLKXHLKHG	173
CYP52A8B	124	GYGIFTLDGEGWKHSRAMLRPQFAREQVAHVTSLEPHFQLLKXHLKHG	173
CYP52D4A	130	GKGIFTLDGPEWKQSRSMRLRPQFAXDRVSHILDLEPHFVLLKXHDGNG	179

* * * * *

Figure 16A

CYP52A1A 383 TLRMYPSVPVNFRTATRDITLPRGGGGANGTDPIYIPKGSTVAYVVYKTHR 432
 CYP52A2A 381 TLRLYPSVPQNFRVATKNTTLPRGGGKDGLSPVLVRKGQTVIYGVYAAHR 430
 CYP52A2B 381 TLRLYPSVPQNFRVATKNTTLPRGGGKDGLSPVLVRKGQTVMYGVYAAHR 430
 CYP52A3A 399 TLRLYPSVPHNFRVATRNITLPRGGGEDGYSPIVVKKGQVVMYTVIATHR 448
 CYP52A3B 399 ALRLYPSVPHNFRVATRNITLPRGGGKDGCSPiVVKKGQVVMYTVIGTHR 448
 CYP52A5A 376 TLRIYPSVPRNFRIATKNTTLPRGGGSDGTSPIILQKGEAVSYGINSTHL 425
 CYP52A5B 376 TLRVYPSVPRNFRIATKNTTLPRGGGPDGTQPIILQKGEVSYGINSTHL 425
 CYP52A8A 371 TLRLHPSVPRNARFAIKDITLPRGGGPNKDPILIRKDEVVQYSISATQT 420
 CYP52A8B 371 TLRLHPSVPRNARFAIKDITLPRGGGPNKDPILIRKNEVVQYSISATQT 420
 CYP52D4A 358 TLRLYPSVPRNARFATRNITLPRGGGPDGSFPILIRKGQPVGYFICATHL 407
 . * * * * . ***** * * . * * * . . .

CYP52A1A 433 LEEYVGKDANDFRPERWFEPSTKKLGWAYVVPFNGGPRVCLGQQFALTEAS 482
 CYP52A2A 431 NPAVYGKDALEFRPERWFEPETKKLGWAFLPFNGGPRICLGQQFALTEAS 480
 CYP52A2B 431 NPAVYGKDALEFRPERWFEPETKKLGWAFLPFNGGPRICLGQQFALTEAS 480
 CYP52A3A 449 DPSIYGADADVFRPERWFEPETRKLWAYVVPFNGGPRICLGQQFALTEAS 498
 CYP52A3B 449 DPSIYGADADVFRPERWFEPETRKLWAYVVPFNGGPRICLGQQFALTEAS 498
 CYP52A5A 426 DPVYGGPDAAEFRPERWFEPSTKKLGWAYLFPFNGGPRICLGQQFALTEAG 475
 CYP52A5B 426 DPVYGGPDAAEFRPERWFEPSTRKLWAYLFPFNGGPRICLGQQFALTEAG 475
 CYP52A8A 421 NPAYYGADAADFRPERWFEPSTRNLGWAFLPFNGGPRICLGQQFALTEAG 470
 CYP52A8B 421 NPAYYGADAADFRPERWFEPSTRNLGWAYLFPFNGGPRICLGQQFALTEAG 470
 CYP52D4A 408 NEKVYGNDSHVFRPERWAALGKSLGWSYLPFNGGPRSCLGQQFAILEAS 457
 * * . ***** . ****...***** ***** * *

CYP52A1A 483 YVITRLAQMFETVSSDPGLEYPKPKCIHLTMSHNDGVFVKM 523
 CYP52A2A 481 YVTVRLQLQEFALHLSMDPDTEYPPKQMSHLTMSLFDGANIEM 522
 CYP52A2B 481 YVTVRLQLQEFHLSMDPDTEYPPKQMSHLTMSLFDGANIEM 522
 CYP52A3A 499 YVTVRLQLQEFALHLSMDPDTEYPPKLQNTLTLSLFDGADVRY 540
 CYP52A3B 499 YVTVRLQLQEFNLSDPNAEYPPKLQNTLTLSLFDGADVRY 540
 CYP52A5A 476 YVLVRLVQEFSHVRLDPDEVYPPKRLTNLTMCLODGAIVKFD 517
 CYP52A5B 476 YVLVRLVQEFSHIRLDPDEVYPPKRLTNLTMCLODGAIVKFD 517
 CYP52A8A 471 YVLVRLVQEFPNLSQDPETKYPPRLAHLTMCLFDGAHVXMS 512
 CYP52A8B 471 YVLVRLVQEFPSLSQDPETKYPPRLAHLTMCLFDGAYVKMQ 512
 CYP52D4A 458 YVLARLTQCYTTIQLR-TTEYPPKKLVHLTMSLLNGVYIRTRT 499
 * * * * . * * * * * * * * * *

Figure 16C

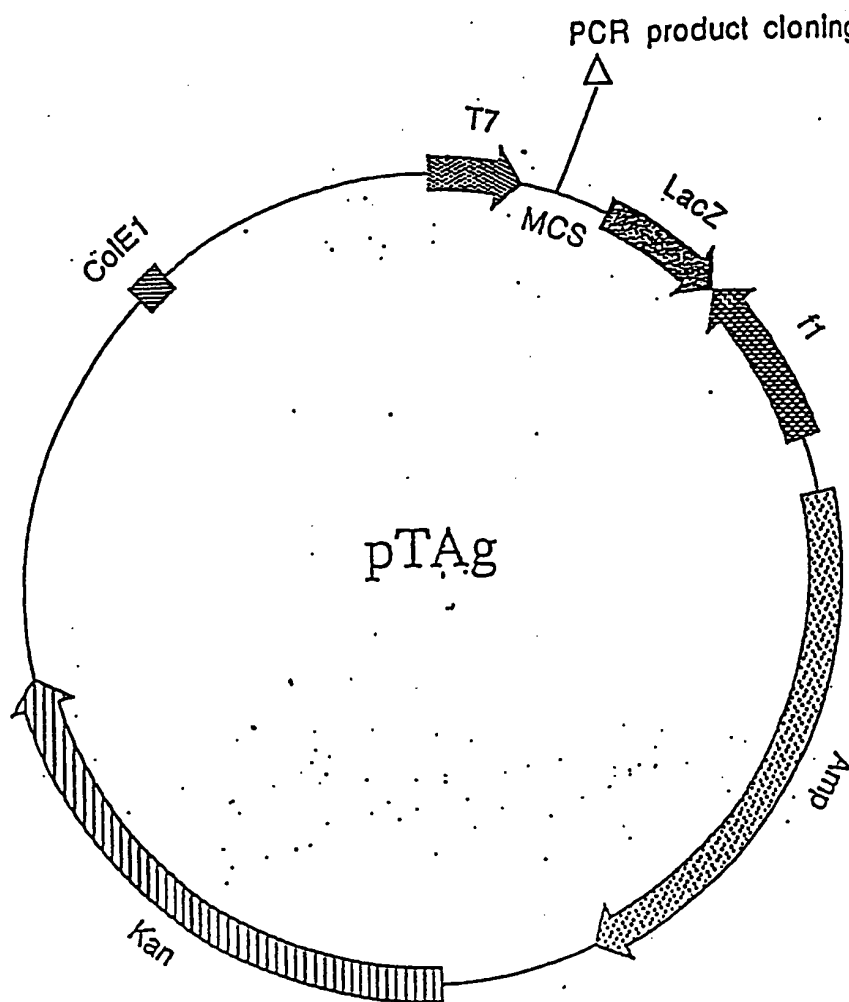
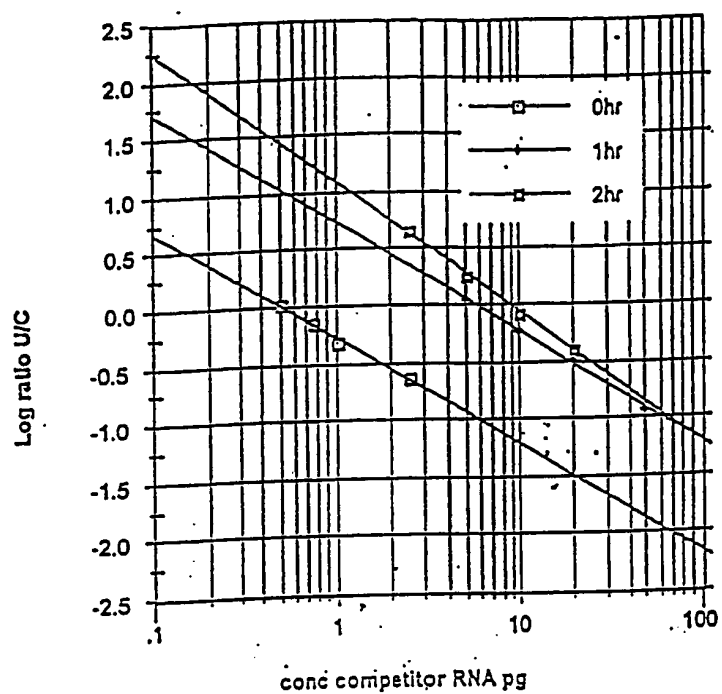
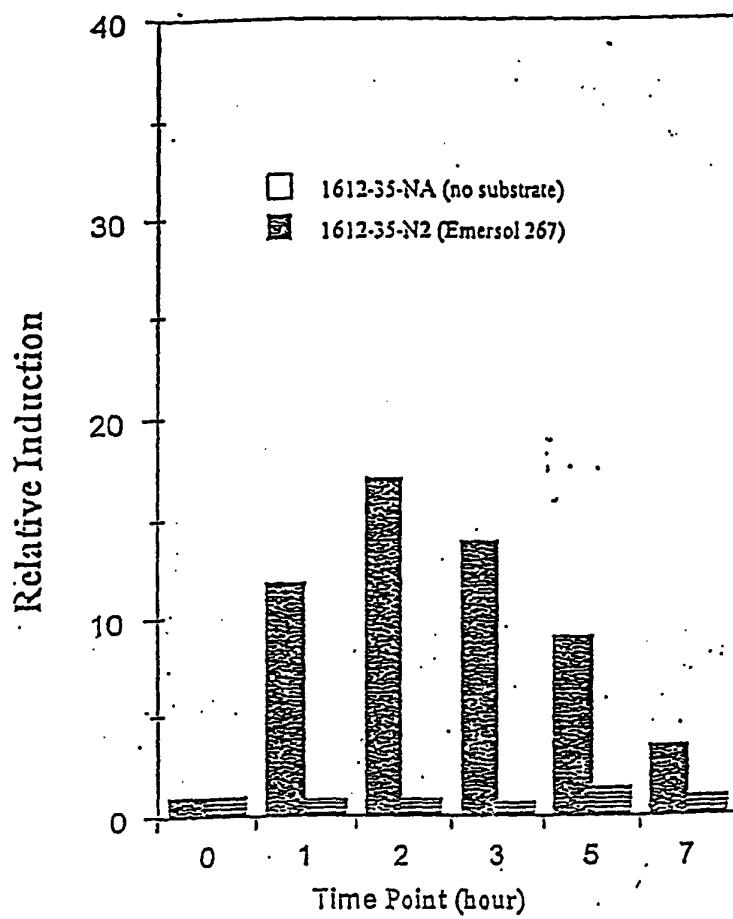
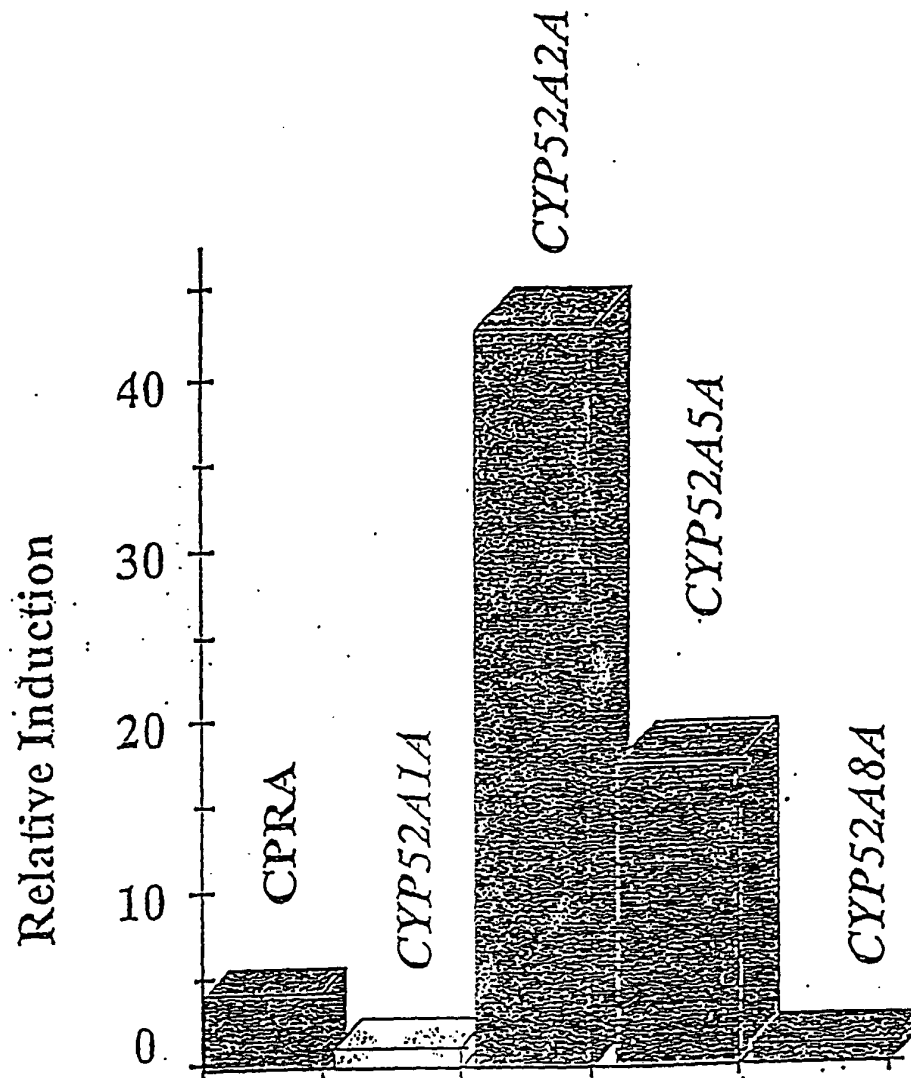
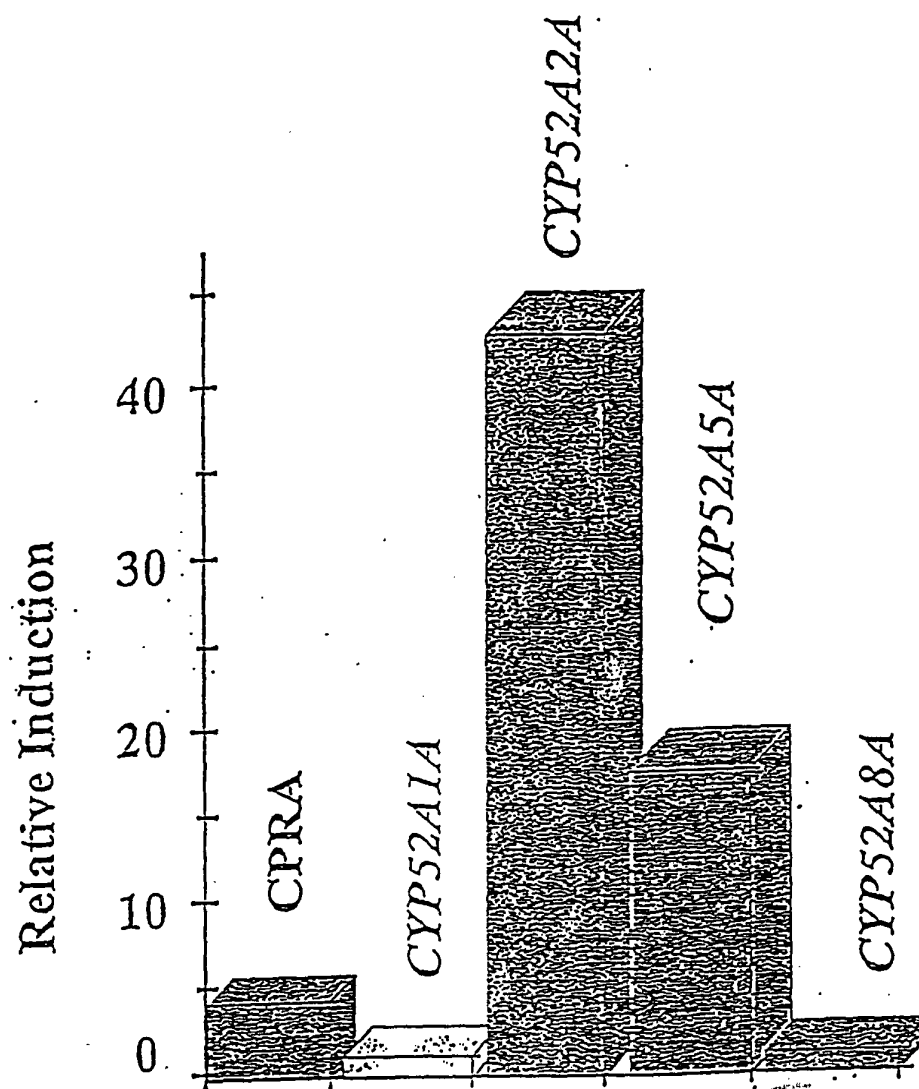


Figure 17

*Figure 18*

*Figure 19*

*Figure 20*

*Figure 20*

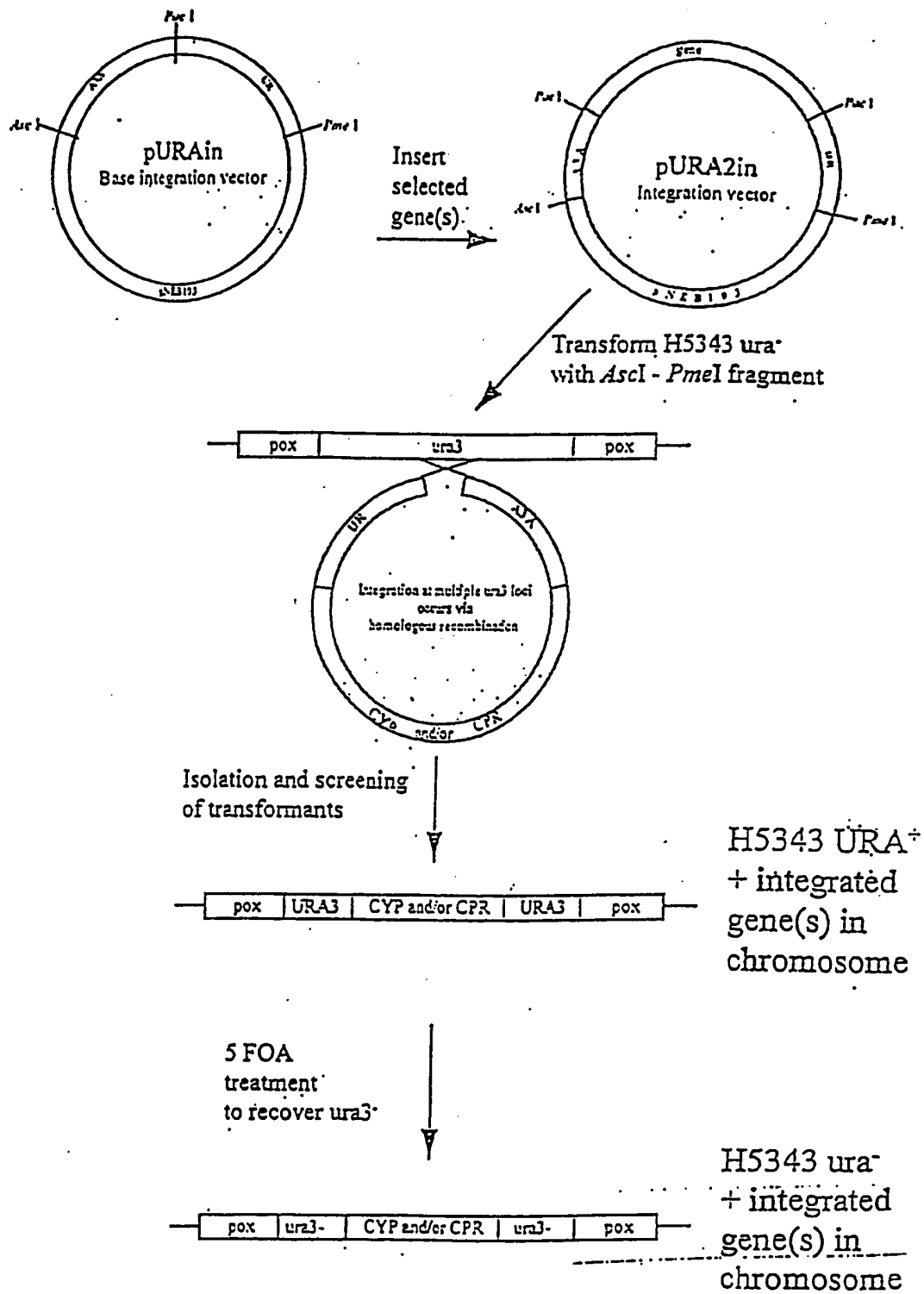
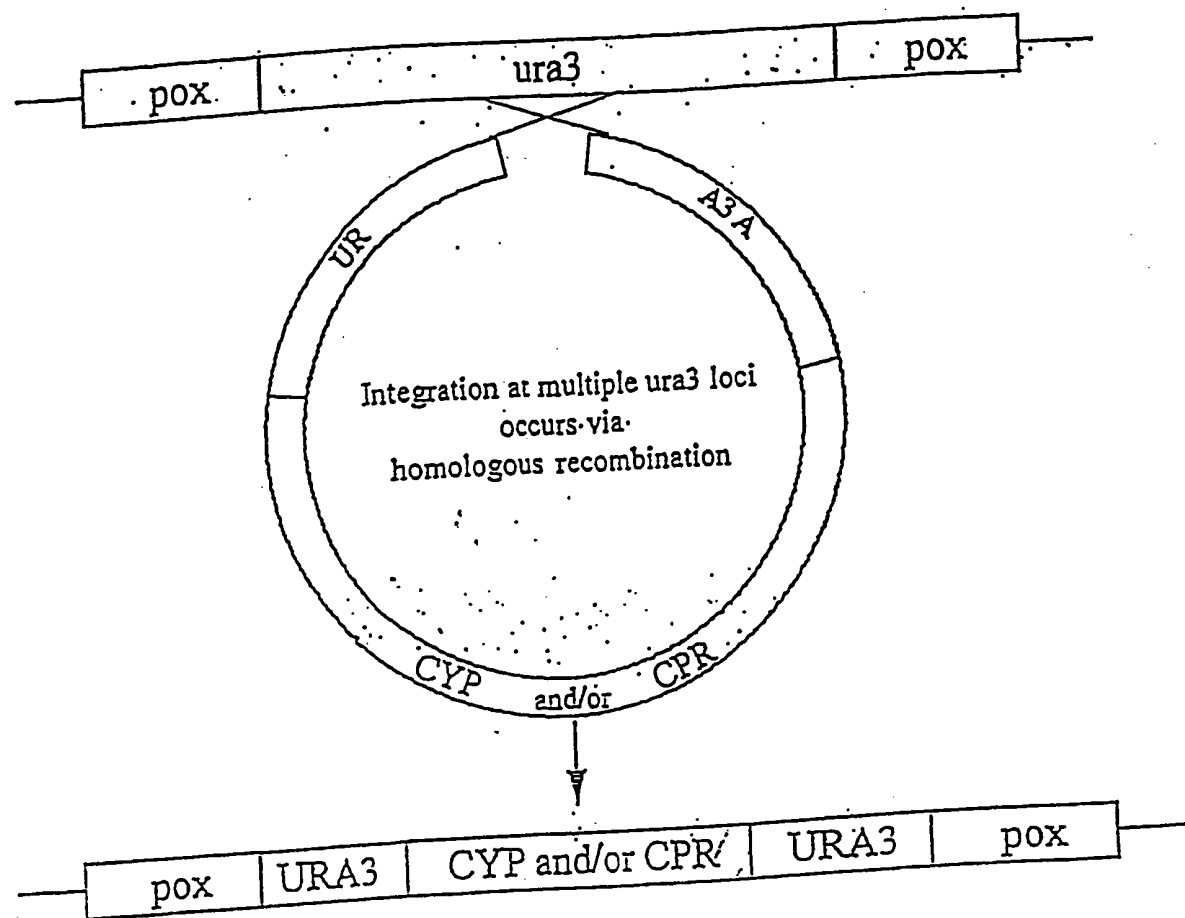


Figure 21

*Figure 22*

Sequence Range: 1 to 1712

10 20 30 40 50 60 70 80
GCTACCGAGC TCACGAGTTC TGGGATTTTC GAGTTTGGAT TTTTTCCTTC GTTCATTGAA TTACGAAAC CAGAGGTTTT
90 100 110 120 130 140 150 160
CAAGACAGAT AAGATTGGGT TTATCAAAAC GCAATTTGAA ATATTCGAGC TGGTTTCCAA GATATCTTGA AGAAGATTGA
170 180 190 200 210 220 230 240
GCAATTTGAA TTGGAAGAG TGGGAGAGAT GTGTTTGGAA TTGTTGGAGA ATTTCAGGAA TGTCAAGATT TACTTTAAGC
250 260 270 280 290 300 310 320
ACGGGTACAA CGAGAATTGT ATTGAATTGA TCAAGAACAT GATCTTGGTG TTACAGAAAC TCAAGTTCTT GCACCAAGAT
330 340 350 360 370 380 390 400
GAGAATGCCA CAGATATACA AGGCGTCATG TGATAAAATG GATGAGATTG ATCCCACTAT TGAAGAAAGA GTTTATGCAA
410 420 430 440 450 460 470 480
AGTGCTCAAC CAGAAGCTAA ACAGGAAGAA GCAAGCAAG AGTGGAAGA AGAAGAAAGA GTTAATTAAG TATTTTGTAT
490 500 510 520 530 540 550 560
TATATAACAA ACAAAGTAAG GAATACAGAT TTATACAATA AATTGCCATA CTAGTCACGT CAGATATCTC ATCCATTCGC
570 580 590 600 610 620 630 640
CAACTCCCAA GAAAAAATAA AAGTGAATAA AAAATGCAAA CCAAAAGATC AACCTCCCAA TCATCATCTG CATCAAAACC
650 660 670 680 690 700 710 720
CCAGCTCAT TCCCAATGCT TACCAACAAA ACATACACAG AAAGGCGATC AGCAGACCCC TCCAGGTTTG CCCAAGCTTT
M V S T K T Y T E R A S A H P S X V A Q R L
730 740 750 760 770 780 790 800
ATTCCGCTTA ATGGAATCCA AAAGAACCAA CCTCTGCGCC TCGATCGAGC TGACCAACAG CCGCGAGTTC GTTTGCTTCA
F R L M E S K K T N L C A S I D V T T A E F L S L
810 820 830 840 850 860 870 880
TCSACAGCT CCGTCCCGAC ATCTGTCTCC TGAAGACCAA CATCGATATC ATCTCAGACT TCAGGTACGA GGGCAAGATT
I D K L G P H I C L V K T H I D I C S D F S Y E G T I
890 900 910 920 930 940 950 960
GAGCGCTTGC TTGTGCTTGC AAAGCGCCAG GCGTTTGTGA TATTGAGGAA CAGGAAGATT GGTGATACCG GAAGACGCT
E P L L V L A E R H G F L I F E D R K F A D I G N T V
970 980 990 1000 1010 1020 1030 1040
GATGTTGAGC TACAGCTCGG GCGTATACCG GATCGCGCGC TCGAGTGAGA TCAGCAACCG GCAAGGAGTG ACTGGAAGG
H L Q Y T S G V Y R I A A H S D I T N A H G V T G K
1050 1060 1070 1080 1090 1100 1110 1120
GCGTCTTTGA AGGTTGAAA CCGGCTGCGC AGCGGTGAGA AAAGCAAGCG CCGGCTTTGA TTTGCGCGGA GTTGTGAGT
G V V E G L K R G A E G V E K E R G V L H L A E L S S
1130 1140 1150 1160 1170 1180 1190 1200
AAAGGCTGCT TGGCGCATGG TGAATATACC CCGAGACGGA TCGAGATTGC GAAGAGTGAT CCGGAGTCTC TGATTGGGTT
K G S L A H G E Y T R E T I E I A K S D R E F V I G F
1210 1220 1230 1240 1250 1260 1270 1280
CATCGCGGAG CCGGACATCG GCGGTAGAGA AGAAGGTTT GATTGGATCA TCATGACGCC TGGTGTGGCG TTGGATGATA
I A Q R D H G G R E E G F D N I I M T P G V G L D D
1290 1300 1310 1320 1330 1340 1350 1360
AAAGCGATGC GTTGGGCCAG CATTATAGGA CTGTTGATGA GGTGTTCTG ACTGCTACCG ATGCGATTAT TGTGCGAGA
K G D A L G Q Q Y R T V D E V V L T G T D V I I V G R
1370 1380 1390 1400 1410 1420 1430 1440
GCGTTGTTG GAAAGGAAG AGCGCTGAG GTGCGGGAAG AGAGATACAG GCAATGCTGA TGGAAAGCAT ACTGGAAGAG
G L F G K G R E P E V E G K R Y R D A G W K A Y L K R
1450 1460 1470 1480 1490 1500 1510 1520
AACTGCTGAG TTAGAATAAA TATTGTAATA AATAGTGTGA TATACATACA CTAAAGCTCT AGGACGTCAT TATAGTCTC
T G Q L E >
1530 1540 1550 1560 1570 1580 1590 1600
GAAGTTGCT GCTAGTTTAS TTTCATGAT TTGGAAGAG AATAACGCAA TCGATGAGC AGGATGCTG GTAGTGGCT
1610 1620 1630 1640 1650 1660 1670 1680
TCCGTACAAA CCGAGAGTAC CCGCTCTGAA ACCAGCTGAC ATTCGCGCTT TCGTTCATCC GCAATGCTG GTTGAAGCTA
1690 1700 1710
TCCAGTACG AGTTGTAATA CAGTTGAGG AA

Figure 23

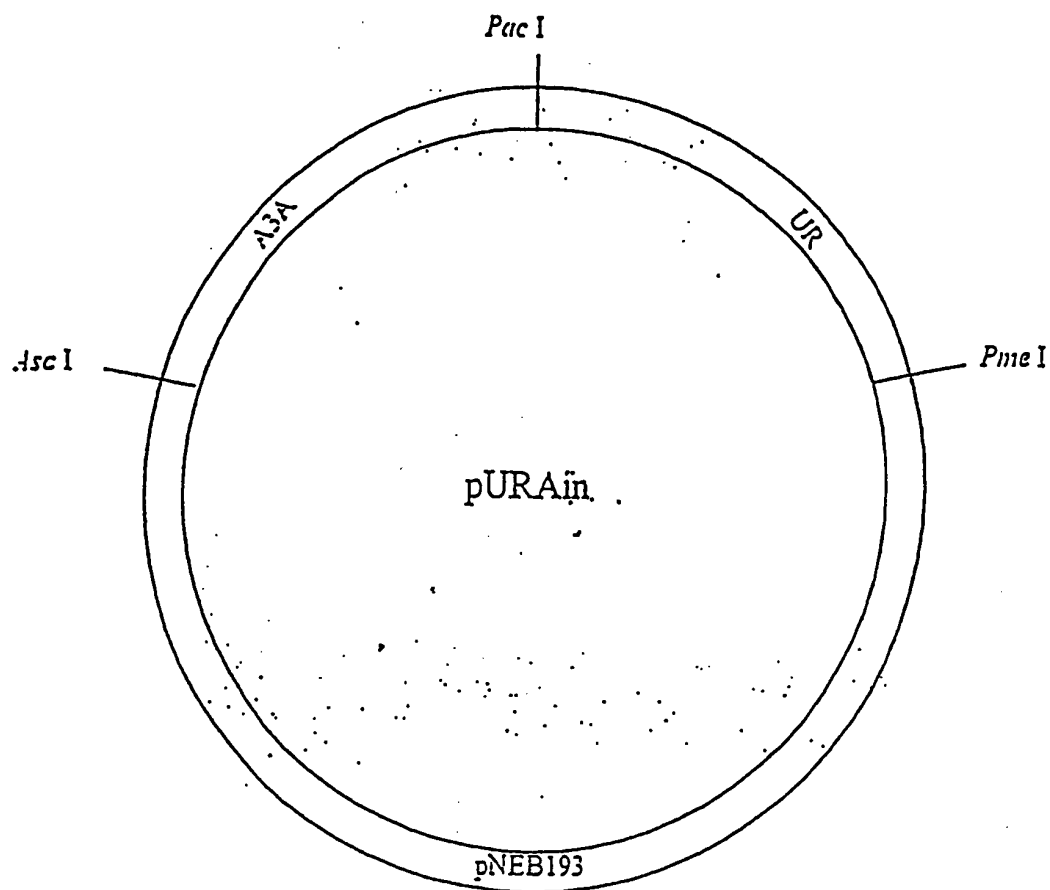
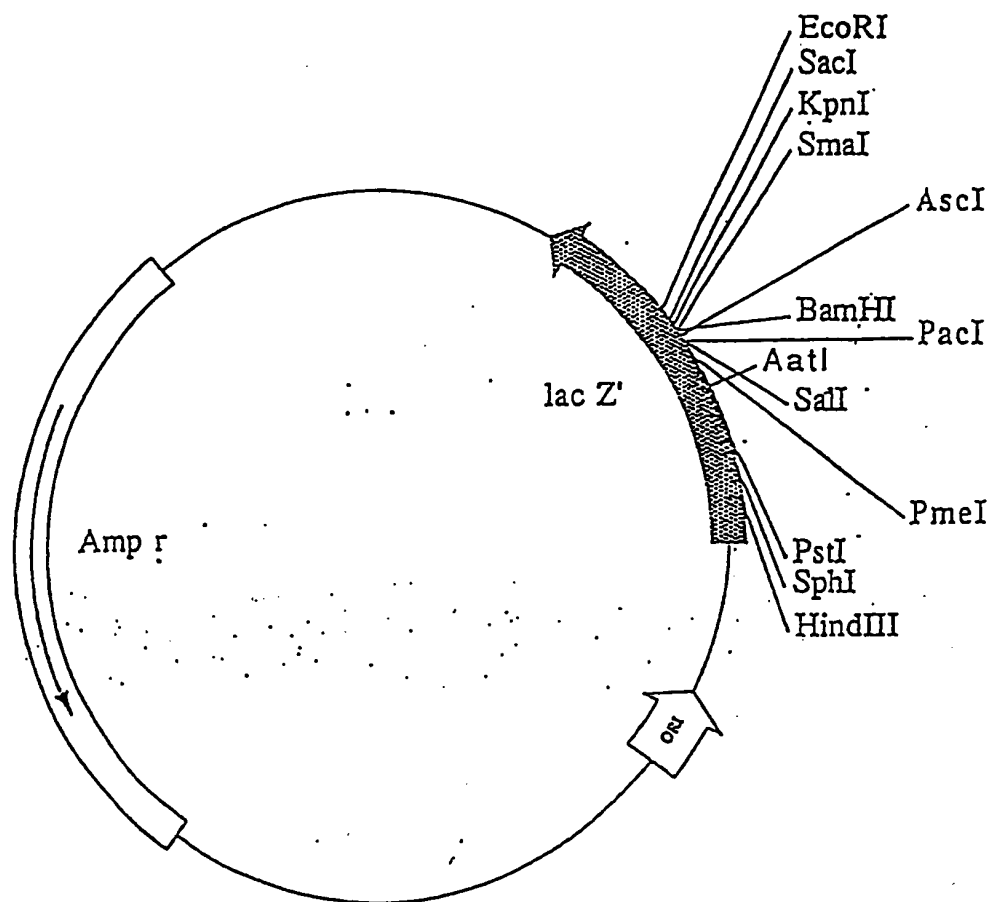


Figure 24

*Figure 25*

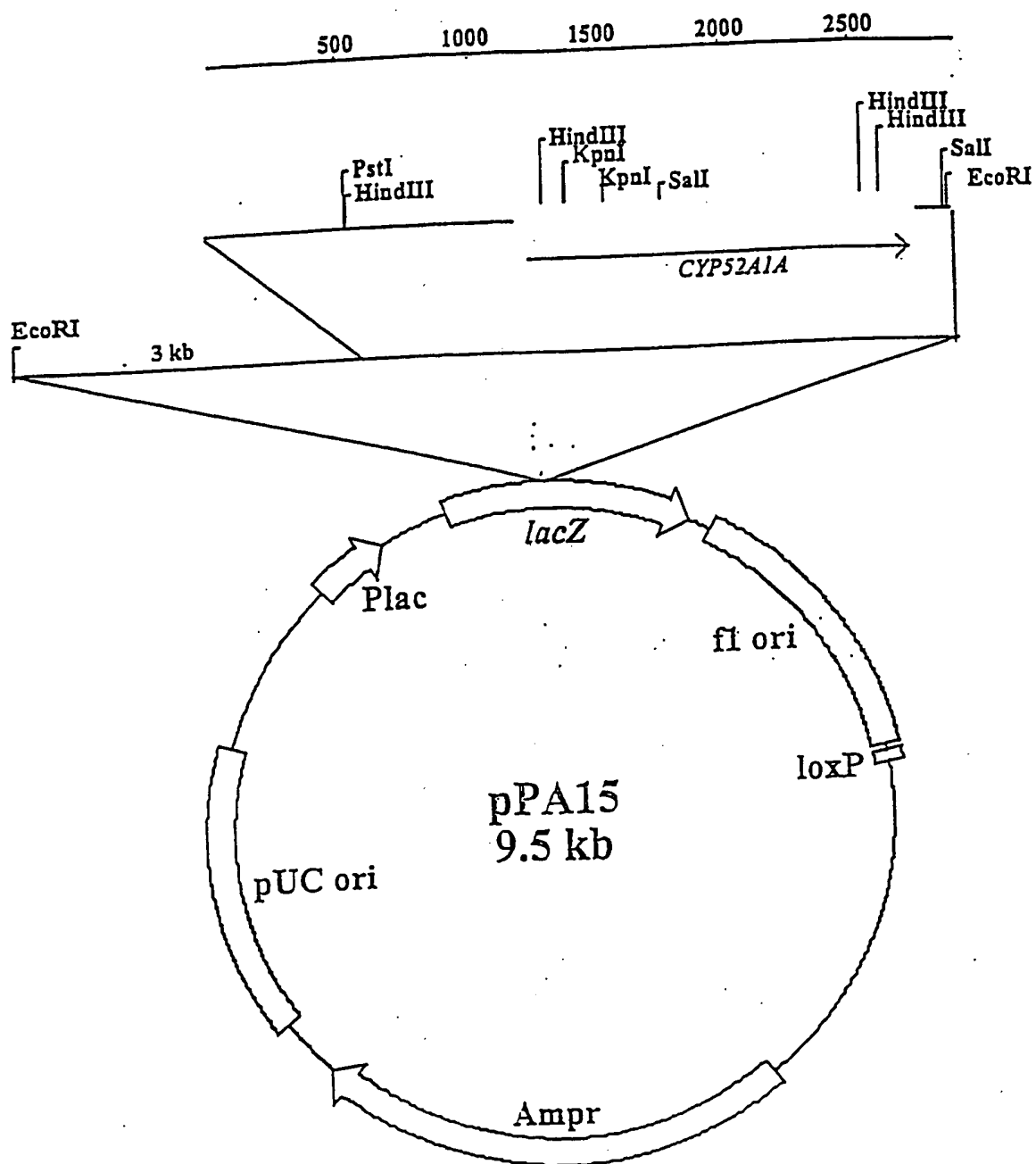


Figure 26

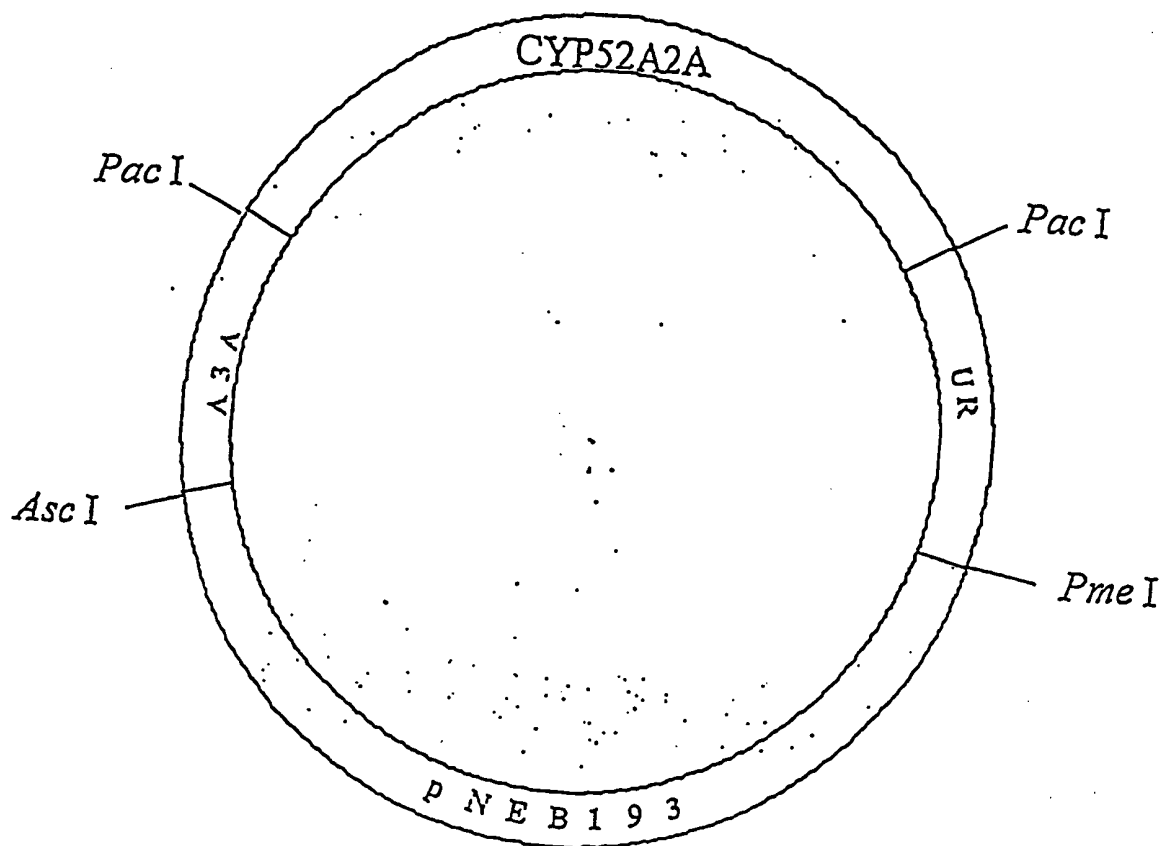


Figure 27

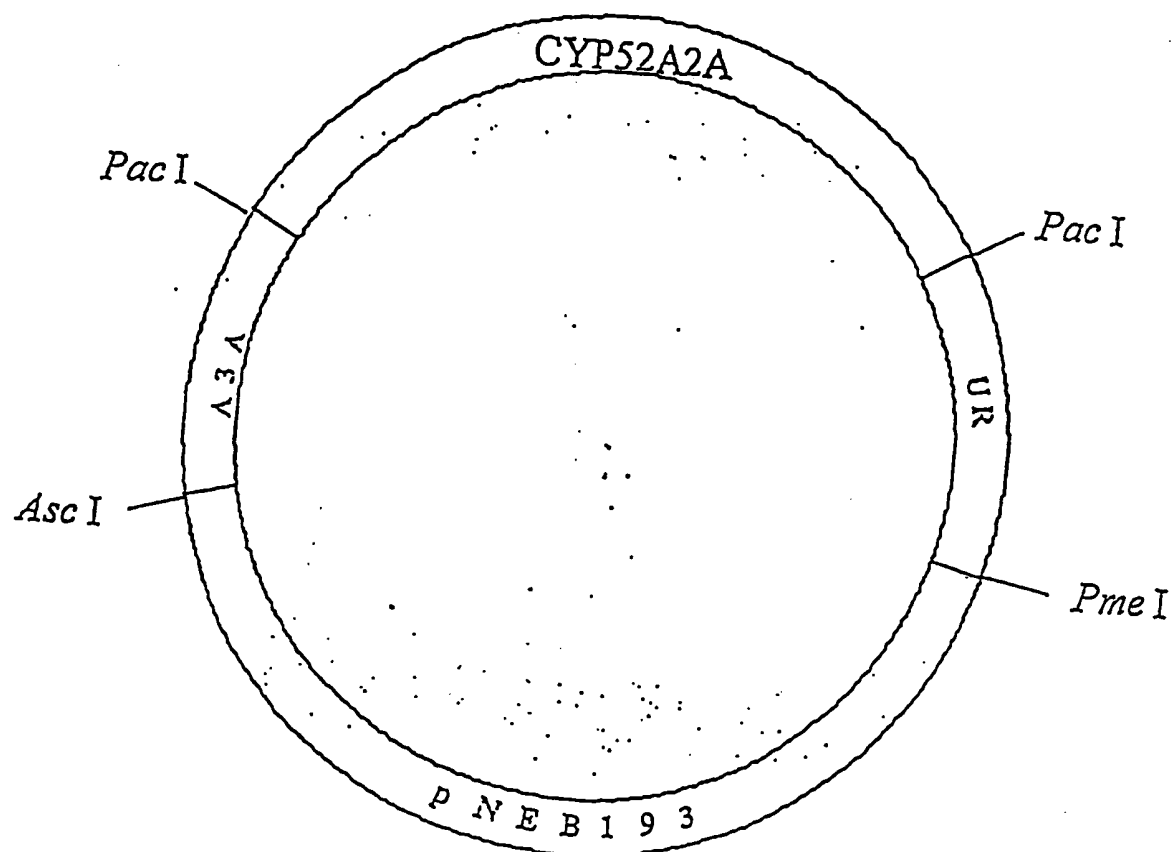
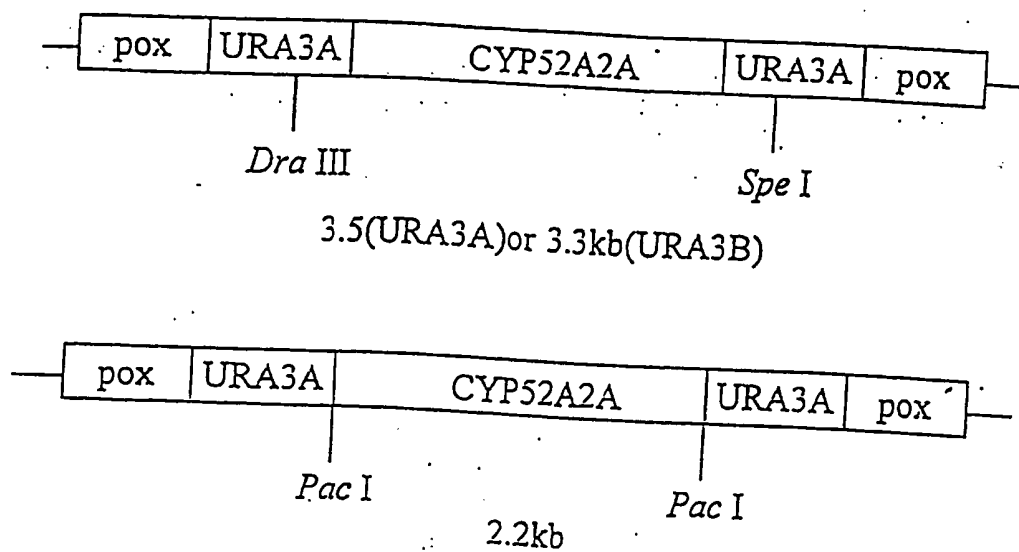


Figure 27

*Figure 28*

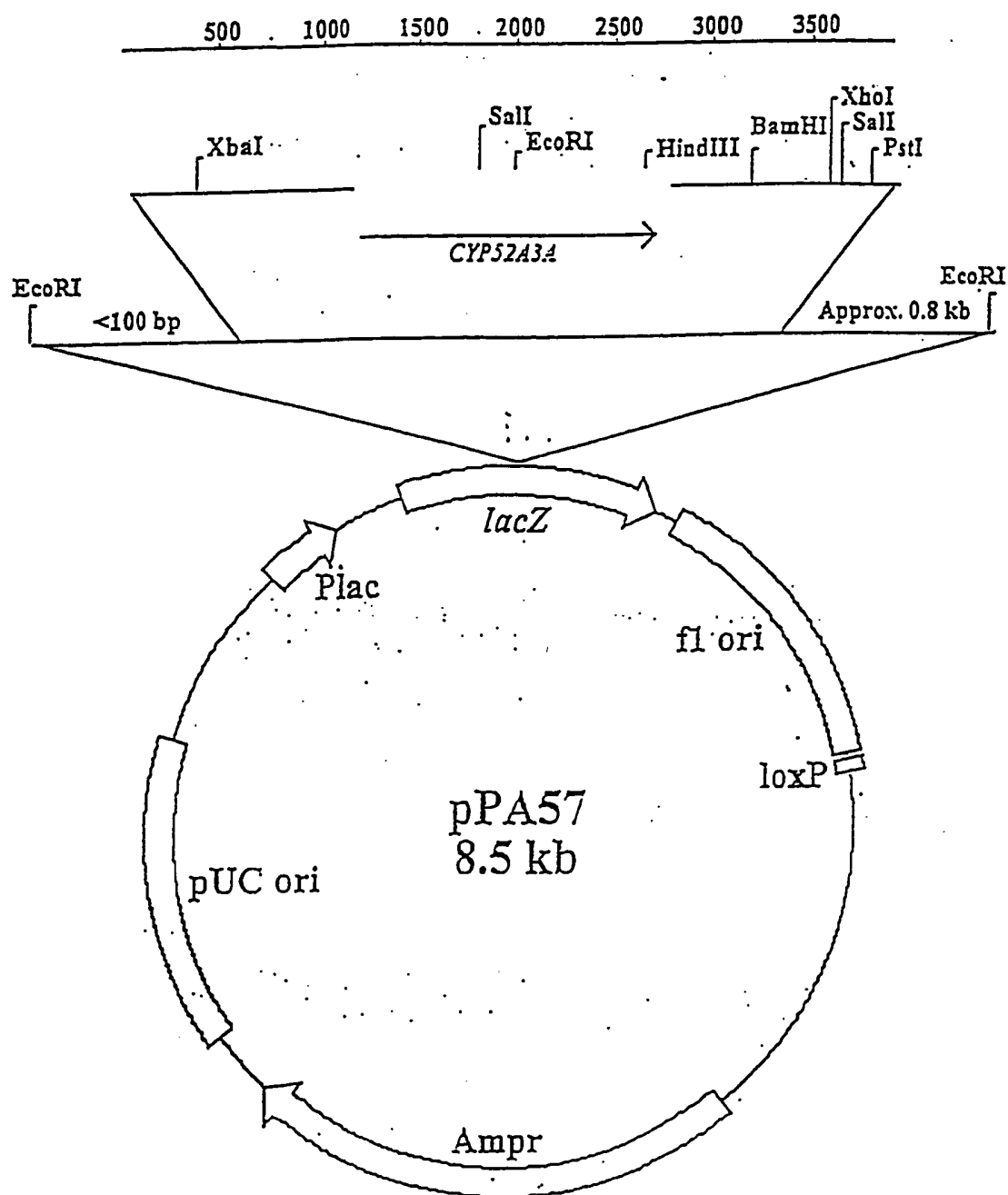


Figure 29

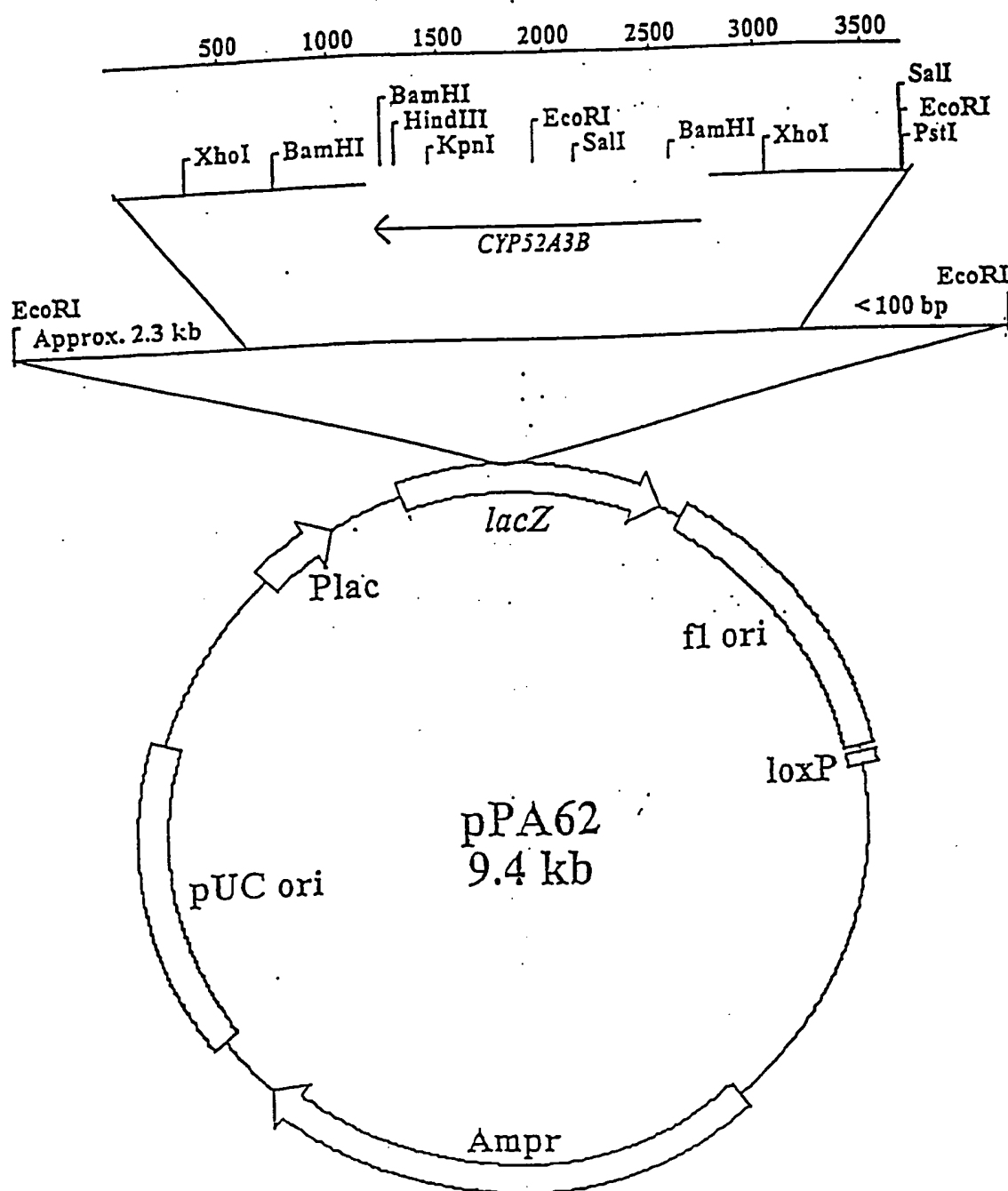
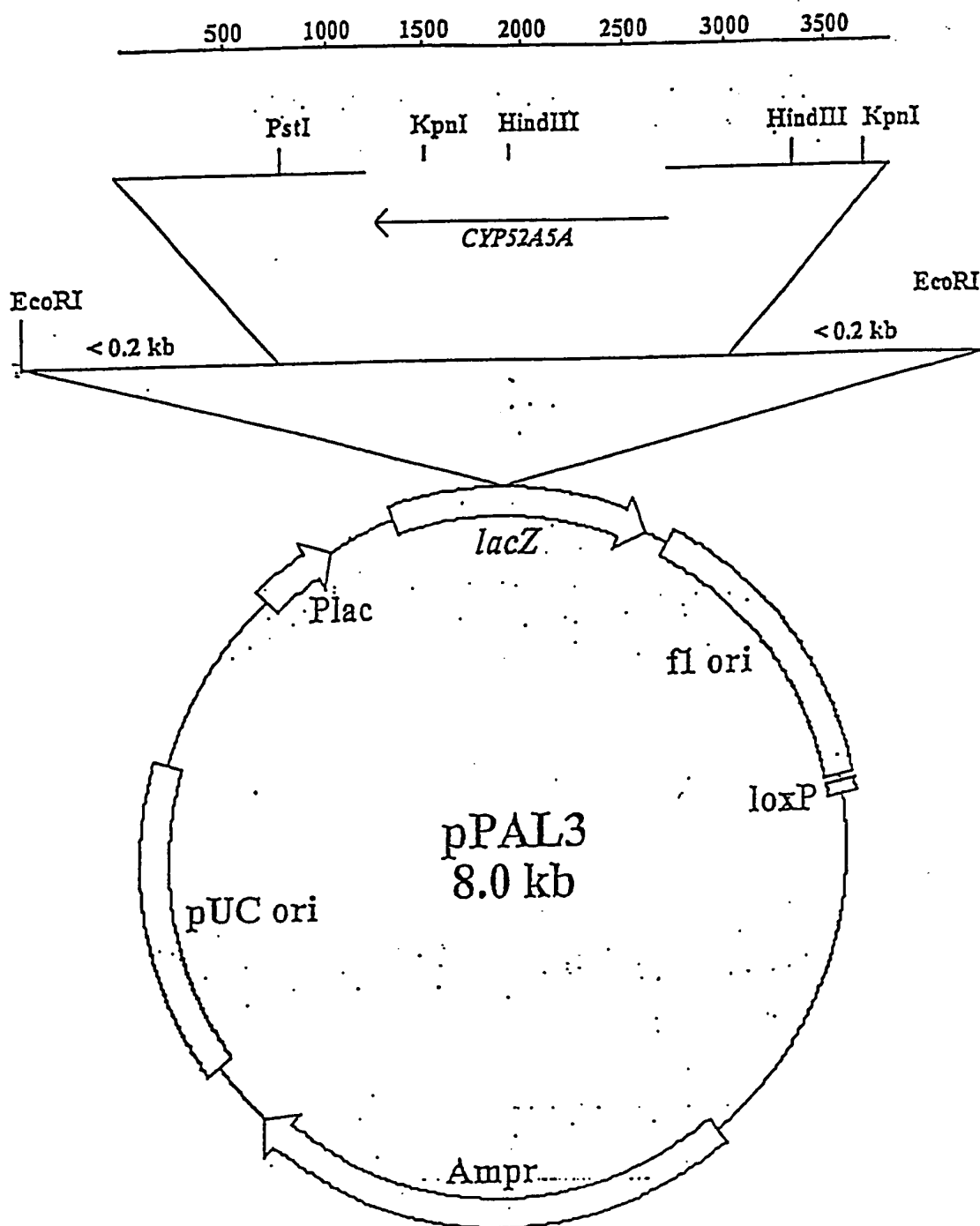


Figure 30

*Figure 31*

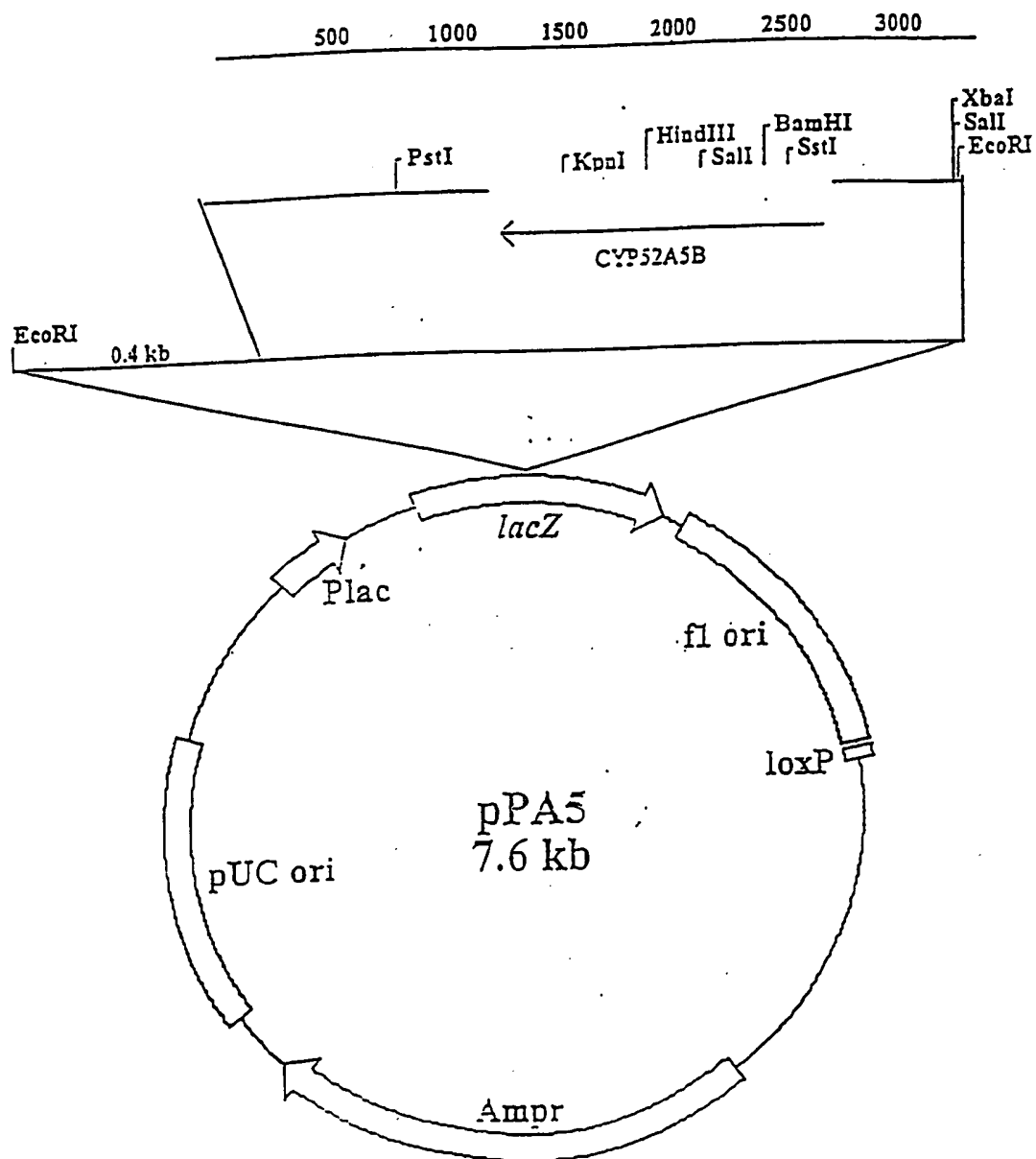
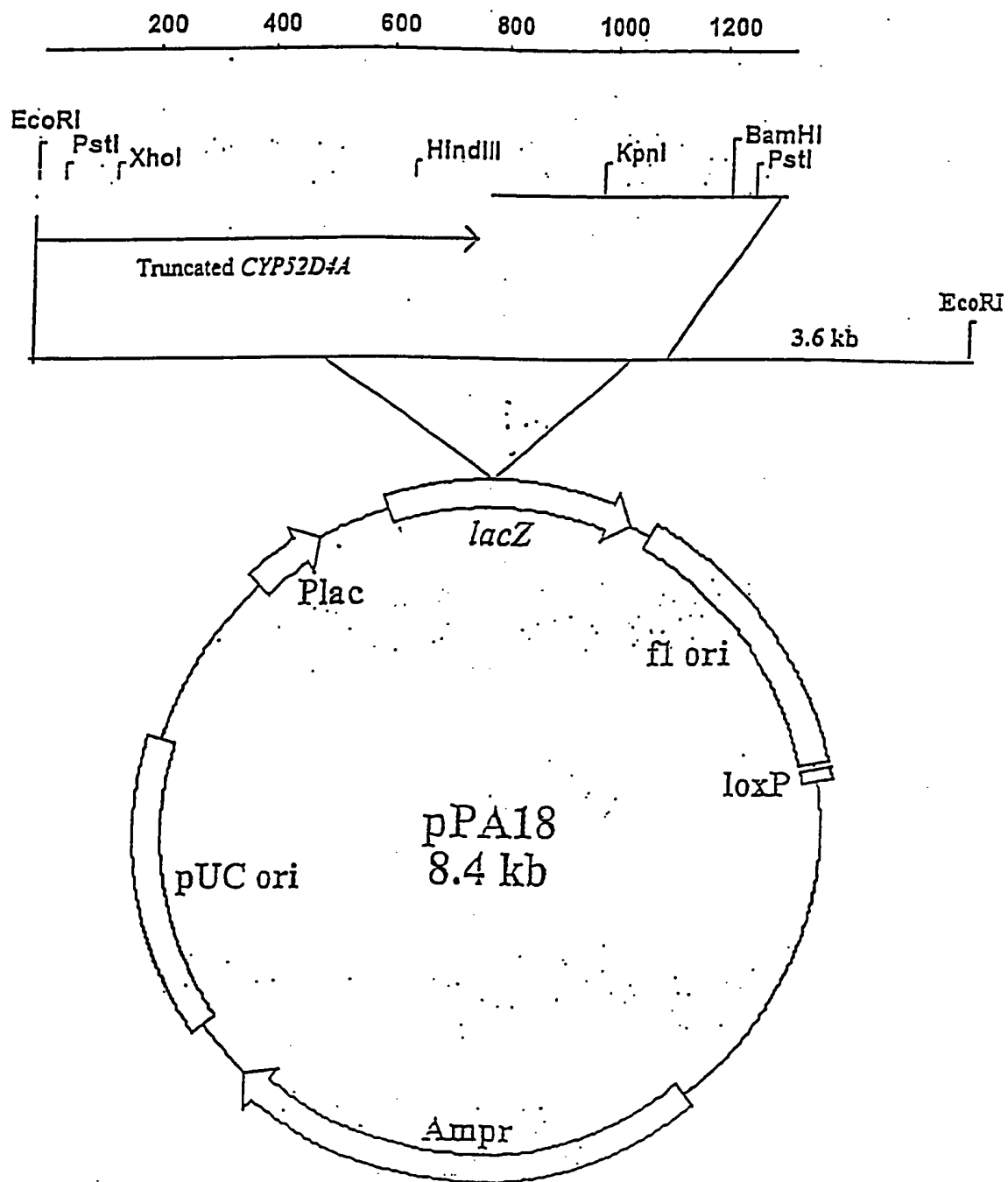


Figure 32

*Figure 33*

Expression of CYP52A1, CYP52A2 and
CYP52A5 in Henkel Fermentor Run 3538-98

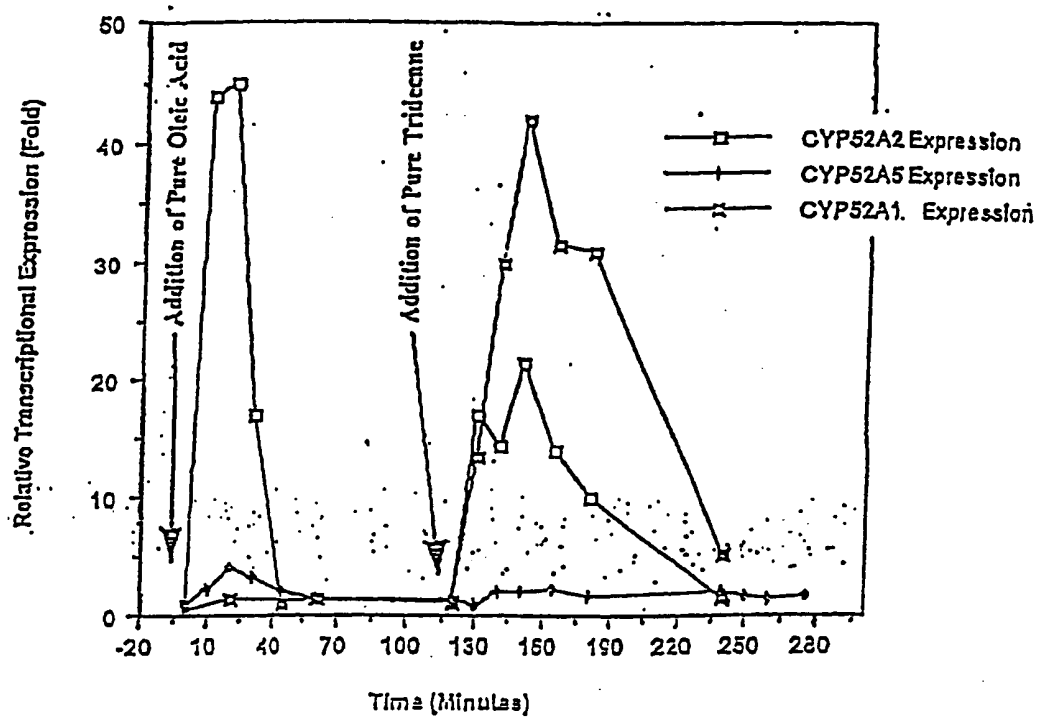
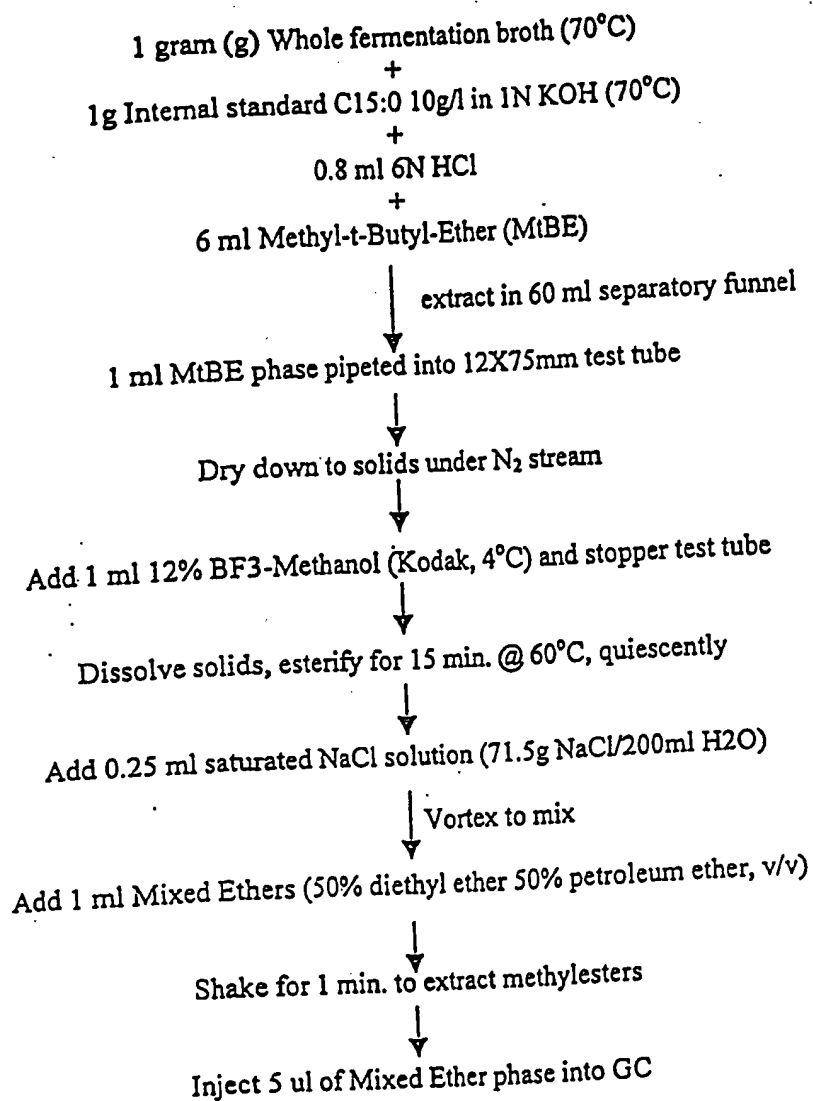


Figure 34

Figure 35**GC parameters**

Column : HP-INNOWAX capillary column, 30m X 0.32mm, 0.5µm film thickness
Split Ratio : 1:100
Column Head Pressure : 13.5 psig
Injector Temperature : 240°C
FID Detector Temp. : 250°C
Temp. Prog. : 90°C for 0 min. to 190°C @ 7°C/min. for 0 min. to 235°C @ 12°C/min. for 30 min.

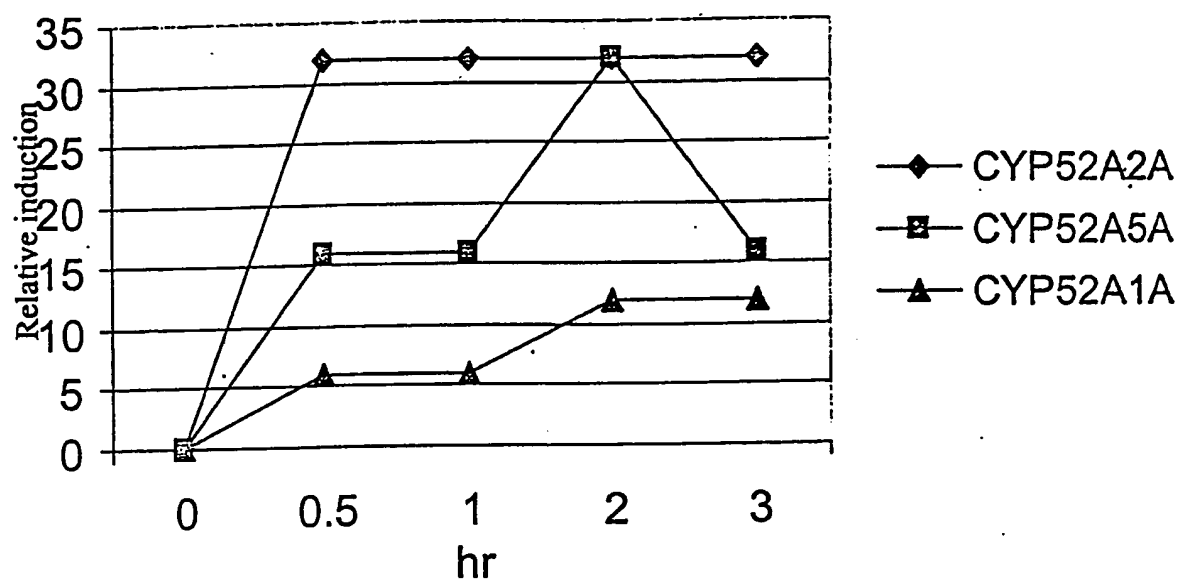


Figure 36

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Wilson, C. Ron
Craft, David L.
Eirich, Dudley
Eshoo, Mark
Madduri, Krishna M.
Cornett, Cathy A.
Brenner, Alfred A.
Tang, Maria
Loper, John C.
Gleeson, Martin
- (ii) TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH
CYTOCHROME P450 OXIDOREDUCTASE GENES AND PROTEINS RELATED
TO THE OMEGA HYDROXYLASE COMPLEX OF *CANDIDA TROPICALIS* AND
METHODS RELATING THERETO
- (iii) NUMBER OF SEQUENCES: 107
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: HENKEL CORPORATION
 - (B) STREET: 2500 Renaissance Boulevard, Suite 200
 - (C) CITY: Gulph Mills
 - (D) STATE: PA
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 19406
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Drach, John E.

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
CCTTAATTAA ATGCACGAAG CGGAGATAAA AG

(2) INFORMATION FOR SEQ ID NO:2:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
CCTTAATTAA GCATAAGCTT GCTCGAGTCT

30

(2) INFORMATION FOR SEQ ID NO:3:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
CCTTAATTAA ACGCAATGGG AACATGGAGT G

31

(2) INFORMATION FOR SEQ ID NO:4:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
CCTTAATTAA TCGCACTACG GTTATTGGTA TCAG

34

(2) INFORMATION FOR SEQ ID NO:5:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
CCTTAATTAA TCAAAGTACG TTCAGGCGG

29

(2) INFORMATION FOR SEQ ID NO:6:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
CCTTAATTAA GGCAGACAAC AACTTGGCAA AGTC

34

(2) INFORMATION FOR SEQ ID NO:7:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGGTTTAAAC

10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGGCGCGCC

9

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCTTAATTAA

10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 3..4
- (D) OTHER INFORMATION: /note= "y=dCTP or dTTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 9..10
- (D) OTHER INFORMATION: /note= "w=dATP or dTTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 15..16
- (D) OTHER INFORMATION: /note= "w=dATP or dTTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 18..19
- (D) OTHER INFORMATION: /note= "w=dATP or dTTP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TCYCAAACWG GTACWGCWGA A

21

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 12..13
- (D) OTHER INFORMATION: /note= "y=dCTP or dTTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 15..16
- (D) OTHER INFORMATION: /note= "w=dATP or dTTP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGTTTGGGTA AYTGWACTTA T

21

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGTTATTATC ATTTCTTC

18

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 3..4
- (D) OTHER INFORMATION: /note= "m=dATP or dCTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 9..10
- (D) OTHER INFORMATION: /note= "r=dATP or dGTP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCMACACCRGTA CCTGGACC

20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATCCCAATCG TAATCAGC

18

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACTTGTCTTC GTTTAGCA

18

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTACGTCTGT GGTGATGC

18

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 3..4
- (D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or dTTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 6..7
- (D) OTHER INFORMATION: /note= "Y=dCTP or dTTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 9..10
- (D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or dTTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 12..13
- (D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or

dTTP"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 15..16
- (D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or

dTTP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGNGAYACNAC NGCNGG

17

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 3..4
(D) OTHER INFORMATION: /note= "r=dATP or dGTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 6..7
(D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 9..10
(D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or
dTTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 12..13
(D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or
dTTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 15..16
(D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or
dTTP"
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
AGRGAYACNA CNGCNGG

17

(2) INFORMATION FOR SEQ ID NO:25:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 3..4
(D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or
dTTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 6..7
(D) OTHER INFORMATION: /note= "r=dATP or dGTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 9..10
(D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 12..13
(D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 15..16

(D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or dTTP"
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:
AGNGCRAAYT GYTGNCC

17

2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 1..2
 - (D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 4..5
 - (D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or dTTP"
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 7..8
 - (D) OTHER INFORMATION: /note= "r=dATP or dGTP"
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 10..11
 - (D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 13..14
 - (D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 16..17
 - (D) OTHER INFORMATION: /note= "n=dATP or dCTP or dGTP or dTTP"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:
YAANGCRAAY TGYTGNCC

18

2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:
ATTCAACGGT GGTCCAAGAA TCTGTTTGG

29

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:
GAGCTATGTT GAGACCACAG TTTGC 25
- (2) INFORMATION FOR SEQ ID NO:29:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:
CTTCAGTTAA AGCAAATTGT TTGGCC 26
- (2) INFORMATION FOR SEQ ID NO:30:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:
CTCGGGAAGC GCGCCATTGT GTTGG 25
- 2) INFORMATION FOR SEQ ID NO:31:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:
TAATACGACT CACTATAGGG CGAATTGGC 29
- (2) INFORMATION FOR SEQ ID NO:32:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 3..4
(D) OTHER INFORMATION: /note= "r=dATP or dGTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 4..5
(D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 16..17
(D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:
TGRYTCAAAC CATCTYTCTG G 21

- (2) INFORMATION FOR SEQ ID NO:33:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:
GGACCGGCGT TAAAGGG

17

- (2) INFORMATION FOR SEQ ID NO:34:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 9..10
 - (D) OTHER INFORMATION: /note= "w=dATP or dTTP"
 - (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 12..13
 - (D) OTHER INFORMATION: /note= "y=dCTP or dTTP"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:
CATAGTCGWA TYATGCTTAG ACC

23

- 2) INFORMATION FOR SEQ ID NO:35:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:
GGACCACCAT TGAATGG

17

- (2) INFORMATION FOR SEQ ID NO:36:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 540 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ATGATTGAAC AACTCCTAGA ATATTGGTAT GTCGTTGTGC CAGTGTTGTA CATCATCAAA	60
CAACTCCTTG CATAACACAAA GACTCGCGTC TTGATGAAAA AGTTGGGTGC TGCTCCAGTC	120
ACAAACAAGT TGTACGACAA CGCTTTCGGT ATCGTCAATG GATGGAAGGC TCTCCAGTTC	180
AAGAAAGAGG GCAGGGCTCA AGAGTACAAC GATTACAAGT TTGACCACTC CAAGAACCCA	240
AGCGTGGGCA CCTACGTCAG TATCTTTTTC GGCACCAGGA TCGTCGTGAC CAAAGATCCA	300
GAGAATATCA AAGCTATTTT GGCAACCCAG TTTGGTGATT TTTCTTTGGG CAAGAGGCAC	360
ACTCTTTTTA AGCCTTTGTT AGGTGATGGG ATCTTCACAT TGGACGGCGA AGGCTGGAAG	420
CACAGCAGAG CCATGTTGAG ACCACAGTTT GCCAGAGAAC AAGTTGCTCA TGTGACGTCG	480
TTGGAACCAC ACTTCCAGTT GTTGAAGAAG CATATTCTTA AGCACAAGGG TGAATACTTT	540

- 2) INFORMATION FOR SEQ ID NO:37:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:
CCGATGAAGT TTTCGACGAG TACCC 25
- 2) INFORMATION FOR SEQ ID NO:38:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:
AAGGCTTTAA CGTGTC CAAT CTGGTC 26
- (2) INFORMATION FOR SEQ ID NO:39:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:
ATTATCGCCA CATACTCAC CAAATGG 27
- (2) INFORMATION FOR SEQ ID NO:40:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:
CGAGATCGTG GATACGCTGG AGTG 24
- (2) INFORMATION FOR SEQ ID NO:41:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
GCCACTCGGT AAC TTTGTCA GGGAC 25
- (2) INFORMATION FOR SEQ ID NO:42:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:
CATTGAACTG AGTAGCCAAA ACAGCC 26
- (2) INFORMATION FOR SEQ ID NO:43:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:
CCTACGTTTG GTATCGCTAC TCCGTTG 27
- (2) INFORMATION FOR SEQ ID NO:44:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:
TTCCAGCCA GCACCGTCCA AG 22
- (2) INFORMATION FOR SEQ ID NO:45:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:
GCAGAGCCGA TCTATGTTGC GTCC 24
- (2) INFORMATION FOR SEQ ID NO:46:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:
TCATTGAATG CTTCCAGGAA CCTCG 25
- (2) INFORMATION FOR SEQ ID NO:47:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:
AAGAGGGCAG GGCTCAAGAG 20
- (2) INFORMATION FOR SEQ ID NO:48:
(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:
TCCATGTGAA GATCCCATCA C 21
- (2) INFORMATION FOR SEQ ID NO:49:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:
CTTGAAGGCC GTGTTGAACG 20
- (2) INFORMATION FOR SEQ ID NO:50:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:
CAGGATTGTG CTGAGTTGCC G 21
- (2) INFORMATION FOR SEQ ID NO:51:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:
CCATTGCCTT GAGATACGCC ATTGGTAG 28
- (2) INFORMATION FOR SEQ ID NO:52:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:
AGCCTTGGTG TCGTTCCTTT CAACGG 26
- (2) INFORMATION FOR SEQ ID NO:53:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:
TTGGGTTTGT TTGTTTCCTG TGCCG 26
- (2) INFORMATION FOR SEQ ID NO:54:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:
CCTTTGACCT TCAATCTGGC GTAGACG 27
- (2) INFORMATION FOR SEQ ID NO:55:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:
GTTTGCTGAA TACGCTGAAG GTGATG 26
- (2) INFORMATION FOR SEQ ID NO:56:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:
TGGAGCTGAA CAACTCTCTC GTCTCGG 27
- (2) INFORMATION FOR SEQ ID NO:57:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:
TTCCTCAACA CGGACAGCGG 20
- (2) INFORMATION FOR SEQ ID NO:58:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:
AGTCAACCAG GTGTGGAACG CGTC 24
- (2) INFORMATION FOR SEQ ID NO:59:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:
GGATCCTAAT ACGACTCACT ATAGGGAGGA AGAGGGCAGG GCTCAAGAG 49
- (2) INFORMATION FOR SEQ ID NO:60:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:
TCCATGTGAA GATCCCATCA CGAGTGTGCC TCTTGCCCAA AG 42
- (2) INFORMATION FOR SEQ ID NO:61:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:
GGATCCTAAT ACGACTCACT ATAGGGAGGC CGATGAAGTT TTCGACGAGT ACCC 54
- (2) INFORMATION FOR SEQ ID NO:62:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:
AAGGCTTTAA CGTGTCCAAT CTGGTCAACA TAGCTCTGGA GTGCTTCCAA CC 52
- (2) INFORMATION FOR SEQ ID NO:63:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:
GGATCCTAAT ACGACTCACT ATAGGGAGGA TTATCGCCAC ATACTTCACC AAATGG 56
- (2) INFORMATION FOR SEQ ID NO:64:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:
CGAGATCGTG GATACGCTGG AGTGCGTCGC TCTTCTCTT CAACAATTCA AG 52

- (2) INFORMATION FOR SEQ ID NO:65:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 49 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:
 CATTGAACTG AGTAGCCAAA ACAGCCCATG GTTTCATCA ATGGGAGGC 49
- (2) INFORMATION FOR SEQ ID NO:66:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 54 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:
 GGATCCTAAT ACGACTCACT ATAGGGAGGG CCACTCGGTA ACTTTGTCAG GGAC 54
- (2) INFORMATION FOR SEQ ID NO:67:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 56 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:
 GGATCCTAAT ACGACTCACT ATAGGGAGGC CTACGTTTGG TATCGCTACT CCGTTG 56
- 2) INFORMATION FOR SEQ ID NO:68:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 48 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:
 TTTCAGCCA GCACCGTCCA AGCAACAAGG AGTACAAGAA ATCGTGTC 48
- (2) INFORMATION FOR SEQ ID NO:69:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 53 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:
 GGATCCTAAT ACGACTCACT ATAGGGAGGG CAGAGCCGAT CTATGTTGCG TCC 53
- (2) INFORMATION FOR SEQ ID NO:70:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 45 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:
TCATTGAATG CTTCCAGGAA CCTCGCCACA TCCATCGAGA ACCGG 45
- (2) INFORMATION FOR SEQ ID NO:71:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:
GGATCCTAAT ACGACTCACT ATAGGGAGGC TTGAAGGCCG TGTTGAACG 49
- (2) INFORMATION FOR SEQ ID NO:72:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:
CAGGATTTGT CTGAGTTGCC GCCTGATCAA GATAGGATCC TTGCCG 46
- (2) INFORMATION FOR SEQ ID NO:73:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:
GGATCCTAAT ACGACTCACT ATAGGGAGGG GTTTGCTGAA TACGCTGAAG GTGATG 56
- (2) INFORMATION FOR SEQ ID NO:74:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:
TGGAGCTGAA CAACTCTCTC GTCTCGGGTG GTCGAATGGA CCCTTGGTCA AG 52
- (2) INFORMATION FOR SEQ ID NO:75:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:
GGATCCTAAT ACGACTCACT ATAGGGAGGT TCCTCAACAC GGACAGCGG 49
- (2) INFORMATION FOR SEQ ID NO:76:
(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

AGTCAACCAG GTGTGGAACG CGTCGGTGGC AACATGAAA AACACCAAG

49

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GGATCCTAAT ACGACTCACT ATAGGGAGGC CATTGCCTTG AGATACGCCA TTGGTAG

57

2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

AGCCTTGGTG TCGTTCTTTT CAACGGAAGG TGGTCTCGAT GGTGTGTTCA ACC

53

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

GGATCCTAAT ACGACTCACT ATAGGGAGGT TGGTTTGTGTT TGTTCCTGT GTCCG

55

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

CCTTTGACCT TCAATCTGGC GTAGACGCAG CACCACCGAT CCACCACTTG

50

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4206 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

CATCAAGATC	ATCTATGGGG	ATAATTACGA	CAGCAACATT	GCAGAAAGAG	CGTTGGTCAC	60
AATCGAAAGA	GCCTATGGCG	TTGCCGTCGT	TGAGGCRAAT	GACAGCACCA	ACAATAACGA	120
TGGTCCCAGT	GAAGAGCCTT	CAGAACAGTC	CATTGTTGAC	GCTTAAGGCA	CGGATAATTA	180
CGTGGGGCAA	AGGAACGCGG	AATTAGTTAT	GGGGGGATCA	AAAGCGGAAG	ATTTGTGTTG	240
CTTGTTGGTT	TTTTCTTTTA	TTTTTCATAT	GATTTCTTTG	CGCAAGTAAC	ATGTGCCAAT	300
TTAGTTTGTG	ATTAGCGTGC	CCCACAATTG	GCATCGTGGG	CGGGCGTGTT	TTGTCATACC	360
CCAAGTCTTA	ACTAGCTCCA	CAGTCTCGAC	GGTGTCTCGA	CGATGTCTTC	TTCCACCCCT	420
CCCATGAATC	ATTCAAAGTT	GTTGGGGGAT	CTCCACCAAG	GGCACCGGAG	TTAATGCTTA	480
TGTTTCTCCC	ACTTTGGTTG	TGATTGGGGT	AGTCTAGTGA	GTTGGAGATT	TTCTTTTTTT	540
CGCAGGTGTC	TCCGATATCG	AAATTTGATG	AATATAGAGA	GAAGCCAGAT	CAGCACAGTA	600
GATTGCCTTT	GTAGTTAGAG	ATGTTGAACA	GCAACTAGTT	GAATTACACG	CCACCACTTG	660
ACAGCAAGTG	CAGTGAGCTG	TAAACGATGC	AGCCAGAGTG	TCACCACCAA	CTGACGTTGG	720
GTGGAGTTGT	TGTTGTTGTT	GTTGGCAGGG	CCATATTGCT	AAACGAAGAC	AAGTAGCACA	780
AAACCCAAGC	TTAAGAACAA	AAATAAAAAA	AATTCATACG	ACAATTCCAA	AGCCATTGAT	840
TTACATAATC	AACAGTAAGA	CAGAAAAAAC	TTTCAACATT	TCAAAGTTCC	CTTTTTCTTA	900
TTACTTCTTT	TTTTTCTTCT	TTCTTCTTTT	CCTTCTGTTT	TTCTTACTTT	ATCAGTCTTT	960
TACTTGTTTT	TGCAATTCCT	CATCCTCCTC	CTACTCCTCC	TCACCATGGC	TTTAGACAAG	1020
TTAGATTTGT	ATGTCATCAT	AACATTGGTG	GTCGCTGTAG	CCGCCTATTT	TGCTAAGAAC	1080
CAGTTCCTTG	ATCAGCCCCA	GGACACCGGG	TTCTCAACA	CGGACAGCGG	AAGCAACTCC	1140
AGAGACGTCT	TGCTGACATT	GAAGAAGAAT	AATAAAAAACA	CGTTGTTGTT	GTTTGGGTCC	1200
CAGACGGGTA	CGGCAGAAGA	TTACGCCAAC	AAATTGTCCA	GAGAATTGCA	CTCCAGATTT	1260
GGCTTGAAAA	CGATGGTTGC	AGATTTGCTG	GATTACGATT	GGGATAACTT	CGGAGATATC	1320
ACCGAAGACA	TCTTGGTGTT	TTTCATTGTT	GCCACCTATG	GTGAGGGTGA	ACCTACCGAT	1380
AATGCCGACG	AGTTCCACAC	CTGGTTGACT	GAAGAAGCTG	ACACTTTGAG	TACCTTGAAA	1440
TACACCGTGT	TCGGGTTGGG	TAACCTCCAG	TACGAGTTCT	TCAATGCCAT	TGGTAGAAAG	1500
TTTGACAGAT	TGTTGAGCGA	GAAAGGTGGT	GACAGGTTTG	CTGAATACGC	TGAAGGTGAT	1560
GACGGTACTG	GCACCTTGGA	CGAAGATTTT	ATGGCCTGGA	AGGACAATGT	CTTTGACGCC	1620
TTGAAGAATG	ATTTGAACTT	TGAAGAAAAG	GAATTGAAGT	ACGAACCAAA	CGTGAAATTG	1680
ACTGAGAGAG	ACGACTTGTC	TGCTGCTGAC	TCCCAAGTTT	CCTTGGGTGA	GCCAAACAAG	1740
AAGTACATCA	ACTCCGAGGG	CATCGACTTG	ACCAAGGGTC	CATTGACCCA	CACCCACCCA	1800
TACTTGGCCA	GAATCACCAG	GACGAGAGAG	TTGTTGAGCT	CCAAGGACAG	ACACTGTATC	1860
CACGTTGAAT	TTGACATTTT	TGAATCGAAC	TTGAAATACA	CCACCGGTGA	CCATCTAGCT	1920
ATCTGGCCAT	CCAACCTCCG	CGAAAAACATT	AAGCAATTG	CCAAGTGTGT	CGGATTGGAA	1980
GATAAACTCG	ACACTGTTAT	TGAATTGAAG	GCGTTGGACT	CCACTTACAC	CATCCCATTG	2040
CCAACCCCAA	TTACCTACGG	TGCTGTCATT	AGACACCATT	TAGAAATCTC	CGGTCCAGTC	2100
TCGAGACAAT	TCTTTTGTG	AATTGCTGGG	TTTGCTCCTG	ATGAAGAAAC	AAAGAAGGCT	2160
TTTACCAGAC	TTGGTGGTGA	CAAGCAAGAA	TTGCCCGCCA	AGGTCACCCG	CAGAAAGTTC	2220
AACATTGCCG	ATGCCCTGTT	ATATTCTCTC	AACAACGCTC	CATGGTCCGA	TGTTCTTTTT	2280
GAATTCCTTA	TTGAAAACGT	TCCACACTTG	ACTCCACGTT	ACTACTCCAT	TTCGTCTTCG	2340
TCATTGAGTG	AAAAGCAACT	CATCAACGTT	ACTGCAGTTG	TTGAAGCCGA	AGAAGAAGCT	2400
GATGCGAGAC	CAGTCACTGG	TGTTGTCACC	AACCTGTTGA	AGAACGTTGA	AATTGTGCAA	2460
AACAAGACTG	GCGAAAAGCC	ACTTGTCCAC	TACGATTTGA	GCGGCCCAAG	AGGCAAGTTC	2520
AACAAGTTCA	AGTTGCCAGT	GCATGTGAGA	AGATCCAACT	TTAAGTTGCC	AAAGAAGTCC	2580
ACCACCCCA	TTATCTTGAT	TGGTCCAGGT	ACTGGTGTG	CCCCATTGAG	AGGTTTTGTC	2640
AGAGAAAGAG	TTCAACAAGT	CAAGAATGGT	GTCAATGTTG	GCAAGACTTT	GTTGTTTTAT	2700
GGTTGCAGAA	ACTCCAACGA	GGACTTTTGT	TACAAGCAAG	AATGGGCCGA	GTACGTTCT	2760
GTTTTGGGTG	AAAACTTTGA	GATGTTCAAT	GCCTTCTCCA	GACAAGACCC	ATCCAAGAAG	2820
GTTTACGTCC	AGGATAAGAT	TTTAGAAAAC	AGCCAACCTG	TGCACGAGTT	GTTGACTGAA	2880
GGTGCCATTA	TCTACGTCTG	TGGTGTGACC	AGTAGAATGG	CTAGAGACGT	GCAGACCACA	2940
ATTTCCAAGA	TTGTTGCTAA	AAGCAGAGAA	ATTAGTGAAG	ACAAGGCTGC	TGAATTGGTC	3000
AAGTCTCGGA	AGGTCCAAAA	TAGATACCAA	GAAGATGTTT	GGTAGACTCA	AACGAATCTC	3060
TCTTTCTCCC	AACGCATTTA	TGAATCTTTA	TTCTCATTGA	AGCTTTACAT	ATGTTCTACA	3120
CTTTATTTTT	TTTTTTTTTT	TTATTATTAT	ATTACGAAAC	ATAGGTCAAC	TATATATACT	3180
TGATTAAATG	TTATAGAAAC	AATAACTATT	ATCTACTCGT	CTACTTCTTT	GGCATTGACA	3240
TCAACATTAC	CGTTCCCAT	ACCGTTGCCG	TTGGCAATGC	CGGGATATTT	AGTACAGTAT	3300

CTCCAATCCG	GATTTGAGCT	ATTGTAGATC	AGCTGCAAGT	CATTCTCCAC	CTTCAACCAG	3360
TACTTATACT	TCATCTTTGA	CTTCAAGTCC	AAGTCATAAA	TATTACAAGT	TAGCAAGAAC	3420
TTCTGGCCAT	CCACGATATA	GACGTTATTC	ACGTTATTAT	GCGACGTATG	GATGTGGTTA	3480
TCCTTATTGA	ACTTCTCAA	CTTCAAAAAC	AACCCACGT	CCCGCAACGT	CATTATCAAC	3540
GACAAGTTCT	GGCTCACGTC	GTCCGAGCTC	GTCAAGTTCT	CAATTAGATC	GTTCTTGTTA	3600
TTGATCTTCT	GGTACTTTCT	CAATTGCTGG	AACACATTGT	CCTCGTTGTT	CAAATAGATC	3660
TTGAACAAC	TTTTCAACGG	GATCAACTTC	TCAATCTGGG	CCAAGATCTC	CGCCGGGATC	3720
TTCAGAAACA	AGTCCTGCAA	CCCCTGGTCG	ATGGTCTCCG	GGTACAACAA	GTCCAAGGGG	3780
CAGAAGTGTC	TAGGCACGTG	TTTCAACTGG	TTCAACGAAC	ATGTTCCGAC	GATGTCGAG	3840
TTATAGTTAT	CGTACAACCA	TTTTGGTTTG	ATTTGCAAAA	TGACGGAGCT	GATGCCATCA	3900
TTCTCCTGGT	TCCTCTCATA	GTACAACGTC	CACCTCTTCG	AGAGGCTCAA	TTCCCTCGTAG	3960
TTCCCGTCCA	AGATATTCGG	CAACAAGAGC	CCGTACCGCT	CACGGAGCAT	CAAGTCGTGG	4020
CCCTGGTTGT	TCAACTTGTT	GATGAAGTCC	GAGGTCAAGA	CAATCAACTG	GATGTCGATG	4080
ATCTGGTGCG	GGAACAAGTT	CTTGCATTTT	AGCTCGATGA	AGTCGTACAA	CTCACACGTC	4140
GAGATATACT	CCTGTTCCCTC	CTTCAAGAGC	CGGATCCGCA	AGAGCTTGTG	CTTCAAGTAG	4200
TCGTTG						4206

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4145 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

TATATGATAT	ATGATATATC	TTCTGTGTA	ATTATTATTC	GTATTCGTTA	ATACTTACTA	60
CATTTTTTTT	TCTTTATTTA	TGAAGAAAAG	GAGAGTTCGT	AAGTTGAGTT	GAGTAGAATA	120
GGCTGTTGTG	CATACGGGGA	GCAGAGGAGA	GTATCCGACG	AGGAGGAACT	GGGTGAAATT	180
TCATCTATGC	TGTTGCGTCC	TGTACTGTAC	TGTAAATCTT	AGATTTCCTA	GAGGTTGTTT	240
TAGCAATAAA	AGTGTTTCAA	GATACAATT	TACAGGCAAG	GGTAAAGGAT	CAACTGATTA	300
GCGGAAGATT	GGTGTGCTCT	GTGGGGTTCT	TTTATTTTTC	ATATGATTTT	TTTGCGCGAG	360
TAACATGTGC	CAATCTAGTT	TATGATTAGC	GTACCTCCAC	AATTGGCATC	TTGGACGGGC	420
GTGTTTTGTC	TTACCCCAAG	CCTTATTTAG	TTCCACAGTC	TCCGACGGTG	CTCGCCGATG	480
TCCTCTCCCA	CCCCTCGCAG	GAATCATTCC	AAGTTGTTGG	GGGATCTCCT	CCGCAGTTTA	540
TGTTTCATGTC	TTTCCCACTT	TGTTGTGAT	TGGGGTAGCG	TAGTGAGTTG	GTGATTTTCT	600
TTTTTCCGAG	GTGTCTCCGA	TATCGAAGTT	TGATGAATAT	AGGAGCCAGA	TCAGCATGGT	660
ATATTGCCTT	TGTAGATAGA	GATGTTGAAC	AACAACATAGC	TGAATTACAC	ACCACCGCTA	720
AACGATGCGC	ACAGGGTGTC	ACCGCCAAC	GACGTTGGGT	GGAGTTGTTG	TTGGCAGGGC	780
CATATTGCTA	AACGAAGAGA	AGTAGCACAA	AACCAAGGT	TAAGAACAAT	TAAAAAAATT	840
CATACGACAA	TTCCACAGCC	ATTACATAA	TCAACAGCGA	CAAATGAGAC	AGAAAAAACT	900
TTCAACATTT	CAAAGTTCCC	TTTTTCCTAT	TACTTCTTTT	TTTCTTTTCT	TCCTTTCATT	960
TCCTTTCCTT	CTGCTTTTAT	TACTTTACCA	GTCTTTTGCT	TGTTTTTGCA	ATTCCTCATC	1020
CTCCTCTCA	CCATGGCTTT	AGACAAGTTA	GATTGTATG	TCATCATAAC	ATTGGTGGTC	1080
GCTGTGGCCG	CCTATTTTGC	TAAGAACCAG	TTCTTGATC	AGCCCCAGGA	CACCGGGTTC	1140
CTCAACACGG	ACAGCGGAAG	CAACTCCAGA	GACGCTTTCG	TGACATTGAA	GAAGAATAAT	1200
AAAAACACGT	TGTTGTTGTT	TGGGTCCCAG	ACCGGTACGG	CAGAAGATTA	CGCCAACAAA	1260
TTGTCAAGAG	AATTGCACTC	CAGATTTGGC	TTGAAAACCA	TGGTTGCAGA	TTTCGCTGAT	1320
TACGATTGGG	ATAACTTCGG	AGATATCACC	GAAGATATCT	TGGTGTTTTT	CATCGTTGCC	1380
ACCTACGGTG	AGGGTGAACC	TACCGACAAT	GCCGACGAGT	TCCACACCTG	GTTGACTGAA	1440
GAAGCTGACA	CTTTGAGTAC	TTTGAGATAT	ACCGTGTTCCG	GGTTGGGTAA	CTCCACCTAC	1500
GAGTTCCTCA	ATGCTATTGG	TAGAAAGTTT	GACAGATTGT	TGAGTGAGAA	AGGTGGTGAC	1560
AGATTGCTG	AATATGCTGA	AGGTGACGAC	GGCACTGGCA	CCTTGGACGA	AGATTTCATG	1620
GCCTGGAAGG	ATAATGTCTT	TGACGCCTTG	AAGAATGACT	TGAACCTTGA	AGAAAAGGAA	1680
TTGAAGTACG	AACCAAACGT	GAAATTGACT	GAGAGAGATG	ACTTGTCTGC	TGCCGACTCC	1740
CAAGTTTCCT	TGGGTGAGCC	AAACAAGAAG	TACATCAACT	CCGAGGGCAT	CGACTTGACC	1800
AAGGGTCCAT	TCGACCACAC	CCACCCATAC	TTGGCCAGGA	TCACCGAGAC	CAGAGAGTTG	1860

```

TTCAGCTCCA AGGAAAGACA CTGTATTCAC GTTGAATTTG ACATTTCTGA ATCGAACTTG 1920
AAATACACCA CCGGTGACCA TCTAGCCATC TGGCCATCCA ACTCCGACGA AAACATCAAG 1980
CAATTTGCCA AGTGTTCGG ATTGGAAGAT AAACCTCGACA CTGTTATTGA ATTGAAGGCA 2040
TTGGACTCCA CTTACACCAT TCCATTCCCA ACTCCAATTA CTTACGGTGC TGTCAATTAGA 2100
CACCATTTAG AAATCTCCGG TCCAGTCTCG AGACAATTCT TTTTGTGCGAT TGCTGGGTTT 2160
GCTCCTGATG AAGAAACAAA GAAGACTTTC ACCAGACTTG GTGGTGACAA ACAAGAATTC 2220
GCCACCAAGG TTACCCGCAG AAAGTTCAAC ATTGCCGATG CCTTGTTATA TTCCTCCAAC 2280
AACACTCCAT GGTCCGATGT TCCTTTTGAG TTCCTTATTG AAAACATCCA ACACTTGACT 2340
CCACGTTACT ACTCCATTTC TTCTTCGTCG TTGAGTGAAA AACAACTCAT CAATGTTACT 2400
GCAGTCGTTG AGGCCGAAGA AGAAGCCGAT GGCAGACCAG TCACTGGTGT TGTACCAAC 2460
TTGTTGAAGA ACATTGAAAT TGCGCAAAAC AAGACTGGCG AAAAGCCACT TGTTCACTAC 2520
GATTTGAGCG GCCCAAGAGG CAAGTTCAAC AAGTTCAAGT TGCCAGTGCA CGTGAGAAGA 2580
TCCAACTTTA AGTTGCCAAA GAACCTCACC ACCCCAGTTA TCTTGATTGG TCCAGGTACT 2640
GGTGTGCGCC CATTGAGAGG TTTCTGTAGA GAAAGAGTTC AACAAAGTCAA GAATGGTGTC 2700
AATGTTGGCA AGACTTTGTT GTTTTATGGT TGCAGAACT CCAACGAGGA CTTTTGTAC 2760
AAGCAAGAAT GGGCCGAGTA CGCTTCTGTT TTGGGTGAAA ACTTTGAGAT GTTCAATGCC 2820
TTCTCTAGAC AAGACCCATC CAAGAAGGTT TACGTCCAGG ATAAGATTTT AGAAAACAGC 2880
CAACTTGTGC ACGAATTGTT GACCGAAGGT GCCATTATCT ACGTCTGTGG TGACGCCAGT 2940
AGAATGGCCA GAGACGTCCA GACCACGATC TCCAAGATTG TTGCCAAAAG CAGAGAAATC 3000
AGTGAAGACA AGGCCGCTGA ATTGGTCAAG TCCTGGAAAAG TCCAAAATAG ATACCAAGAA 3060
GATGTTTGGT AGACTCAAAC GAATCTCTCT TTCTCCCAAC GCATTTATGA ATATTCTCAT 3120
TGAAGTTTTA CATATGTTCT ATATTTCATT TTTTTTTTAT TATATTACGA AACATAGGTC 3180
AACTATATAT ACTTGATTAA ATGTTATAGA AACAATAATT ATTATCTACT CGTCTACTTC 3240
TTTGCCATTG GCATTGGCAT TGGCATTGGC ATTGCCGTTG CCGTTGGTAA TGCCGGGATA 3300
TTTAGTACAG TATCTCCAAT CCGGATTGTA GCTATTGTAA ATCAGCTGCA AGTCATTCTC 3360
CACCTTCAAC CAGTACTTAT ACTTCATCTT TGACTTCAAG TCCAAGTCAT AAATATTACA 3420
AGTTAGCAAG AACTTCTGGC CATCCACAAT ATAGACGTTA TTCACGTTAT TATGCGACGT 3480
ATGGATATGG TTATCCTTAT TGAACCTTCT AAACCTCAAA AACAAACCCCA CGTCCCGCAA 3540
CGTCATTATC AACGACAAGT TCTGACTCAC GTCGTCGGAG CTCGTCAAGT TCTCAATTAG 3600
ATCGTTCTTG TTATTGATCT TCTGGTACTT TCTCAACTGC TGGAACACAT TGTCTCGTT 3660
GTTCAAATAG ATCTTGAACA ACTTCTTCAA GGGAAATCAAC TTTTCGATCT GGGCCAAGAT 3720
TTCCGCCGGG ATCTTCAGAA ACAAGTCCTG CAACCCCTGG TCGATGGTCT CGGGGTACAA 3780
CAAGTCTAAG GGGCAGAAGT GTCTAGGCAC GTGTTTCAAC TGGTTCAAGG AACATGTTCTG 3840
ACAGTAGTTC GAGTTATAGT TATCGTACAA CCACCTTGGC TTGATTTTGA AAATGACGGA 3900
GCTGATCCCA TCATTCTCCT GGTTCCTTTC ATAGTACAAC TGGCATTCTC TCGAGAGACT 3960
CAACTCCTCG TAGTTCCCGT CCAAGATATT CGGCAACAAG AGCCCGTAGC GCTCAGGAG 4020
CATCAAGTCG TGGCCCTGGT TGTTCAACTT GTTGATGAAG TCCGATGTCA AGACAATCAA 4080
CTGGATGTCG ATGATCTGGT GCGGAAACAA GTTCTTGCAC TTTAGCTCGA TGAAGTCGTA 4140
CAACT 4145

```

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 679 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

```

Met Ala Leu Asp Lys Leu Asp Leu Tyr Val Ile Ile Thr Leu Val Val
1           5           10          15
Ala Val Ala Ala Tyr Phe Ala Lys Asn Gln Phe Leu Asp Gln Pro Gln
20          25          30
Asp Thr Gly Phe Leu Asn Thr Asp Ser Gly Ser Asn Ser Arg Asp Val
35          40          45
Leu Leu Thr Leu Lys Lys Asn Asn Lys Asn Thr Leu Leu Leu Phe Gly
50          55          60

```

Ser Gln Thr Gly Thr Ala Glu Asp Tyr Ala Asn Lys Leu Ser Arg Glu
 65 70 75 80
 Leu His Ser Arg Phe Gly Leu Lys Thr Met Val Ala Asp Phe Ala Asp
 85 90 95
 Tyr Asp Trp Asp Asn Phe Gly Asp Ile Thr Glu Asp Ile Leu Val Phe
 100 105 110
 Phe Ile Val Ala Thr Tyr Gly Glu Gly Glu Pro Thr Asp Asn Ala Asp
 115 120 125
 Glu Phe His Thr Trp Leu Thr Glu Glu Ala Asp Thr Leu Ser Thr Leu
 130 135 140
 Lys Tyr Thr Val Phe Gly Leu Gly Asn Ser Thr Tyr Glu Phe Phe Asn
 145 150 155 160
 Ala Ile Gly Arg Lys Phe Asp Arg Leu Leu Ser Glu Lys Gly Gly Asp
 165 170 175
 Arg Phe Ala Glu Tyr Ala Glu Gly Asp Asp Gly Thr Gly Thr Leu Asp
 180 185 190
 Glu Asp Phe Met Ala Trp Lys Asp Asn Val Phe Asp Ala Leu Lys Asn
 195 200 205
 Asp Leu Asn Phe Glu Glu Lys Glu Leu Lys Tyr Glu Pro Asn Val Lys
 210 215 220
 Leu Thr Glu Arg Asp Asp Leu Ser Ala Ala Asp Ser Gln Val Ser Leu
 225 230 235 240
 Gly Glu Pro Asn Lys Lys Tyr Ile Asn Ser Glu Gly Ile Asp Leu Thr
 245 250 255
 Lys Gly Pro Phe Asp His Thr His Pro Tyr Leu Ala Arg Ile Thr Glu
 260 265 270
 Thr Arg Glu Leu Phe Ser Ser Lys Asp Arg His Cys Ile His Val Glu
 275 280 285
 Phe Asp Ile Ser Glu Ser Asn Leu Lys Tyr Thr Thr Gly Asp His Leu
 290 295 300
 Ala Ile Trp Pro Ser Asn Ser Asp Glu Asn Ile Lys Gln Phe Ala Lys
 305 310 315 320
 Cys Phe Gly Leu Glu Asp Lys Leu Asp Thr Val Ile Glu Leu Lys Ala
 325 330 335
 Leu Asp Ser Thr Tyr Thr Ile Pro Phe Pro Thr Pro Ile Thr Tyr Gly
 340 345 350
 Ala Val Ile Arg His His Leu Glu Ile Ser Gly Pro Val Ser Arg Gln
 355 360 365
 Phe Phe Leu Ser Ile Ala Gly Phe Ala Pro Asp Glu Glu Thr Lys Lys
 370 375 380
 Ala Phe Thr Arg Leu Gly Gly Asp Lys Gln Glu Phe Ala Ala Lys Val
 385 390 395 400
 Thr Arg Arg Lys Phe Asn Ile Ala Asp Ala Leu Leu Tyr Ser Ser Asn
 405 410 415
 Asn Ala Pro Trp Ser Asp Val Pro Phe Glu Phe Leu Ile Glu Asn Val
 420 425 430
 Pro His Leu Thr Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Ser Leu Ser
 435 440 445
 Glu Lys Gln Leu Ile Asn Val Thr Ala Val Val Glu Ala Glu Glu Glu
 450 455 460
 Ala Asp Gly Arg Pro Val Thr Gly Val Val Thr Asn Leu Leu Lys Asn
 465 470 475 480
 Val Glu Ile Val Gln Asn Lys Thr Gly Glu Lys Pro Leu Val His Tyr
 485 490 495
 Asp Leu Ser Gly Pro Arg Gly Lys Phe Asn Lys Phe Lys Leu Pro Val
 500 505 510

His Val Arg Arg Ser Asn Phe Lys Leu Pro Lys Asn Ser Thr Thr Pro
 515 520 525
 Val Ile Leu Ile Gly Pro Gly Thr Gly Val Ala Pro Leu Arg Gly Phe
 530 535 540
 Val Arg Glu Arg Val Gln Gln Val Lys Asn Gly Val Asn Val Gly Lys
 545 550 555 560
 Thr Leu Leu Phe Tyr Gly Cys Arg Asn Ser Asn Glu Asp Phe Leu Tyr
 565 570 575
 Lys Gln Glu Trp Ala Glu Tyr Ala Ser Val Leu Gly Glu Asn Phe Glu
 580 585 590
 Met Phe Asn Ala Phe Ser Arg Gln Asp Pro Ser Lys Lys Val Tyr Val
 595 600 605
 Gln Asp Lys Ile Leu Glu Asn Ser Gln Leu Val His Glu Leu Leu Thr
 610 615 620
 Glu Gly Ala Ile Ile Tyr Val Cys Gly Asp Ala Ser Arg Met Ala Arg
 625 630 635 640
 Asp Val Gln Thr Thr Ile Ser Lys Ile Val Ala Lys Ser Arg Glu Ile
 645 650 655
 Ser Glu Asp Lys Ala Ala Glu Leu Val Lys Ser Trp Lys Val Gln Asn
 660 665 670
 Arg Tyr Gln Glu Asp Val Trp
 675

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 679 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Met Ala Leu Asp Lys Leu Asp Leu Tyr Val Ile Ile Thr Leu Val Val
 1 5 10 15
 Ala Val Ala Ala Tyr Phe Ala Lys Asn Gln Phe Leu Asp Gln Pro Gln
 20 25 30
 Asp Thr Gly Phe Leu Asn Thr Asp Ser Gly Ser Asn Ser Arg Asp Val
 35 40 45
 Leu Leu Thr Leu Lys Lys Asn Asn Lys Asn Thr Leu Leu Phe Gly
 50 55 60
 Ser Gln Thr Gly Thr Ala Glu Asp Tyr Ala Asn Lys Leu Ser Arg Glu
 65 70 75 80
 Leu His Ser Arg Phe Gly Leu Lys Thr Met Val Ala Asp Phe Ala Asp
 85 90 95
 Tyr Asp Trp Asp Asn Phe Gly Asp Ile Thr Glu Asp Ile Leu Val Phe
 100 105 110
 Phe Ile Val Ala Thr Tyr Gly Glu Gly Glu Pro Thr Asp Asn Ala Asp
 115 120 125
 Glu Phe His Thr Trp Leu Thr Glu Glu Ala Asp Thr Leu Ser Thr Leu
 130 135 140
 Arg Tyr Thr Val Phe Gly Leu Gly Asn Ser Thr Tyr Glu Phe Phe Asn
 145 150 155 160
 Ala Ile Gly Arg Lys Phe Asp Arg Leu Leu Ser Glu Lys Gly Gly Asp
 165 170 175
 Arg Phe Ala Glu Tyr Ala Glu Gly Asp Asp Gly Thr Gly Thr Leu Asp
 180 185 190

Glu Asp Phe Met Ala Trp Lys Asp Asn Val Phe Asp Ala Leu Lys Asn
 195 200 205
 Asp Leu Asn Phe Glu Glu Lys Glu Leu Lys Tyr Glu Pro Asn Val Lys
 210 215 220
 Leu Thr Glu Arg Asp Asp Leu Ser Ala Ala Asp Ser Gln Val Ser Leu
 225 230 235 240
 Gly Glu Pro Asn Lys Lys Tyr Ile Asn Ser Glu Gly Ile Asp Leu Thr
 245 250 255
 Lys Gly Pro Phe Asp His Thr His Pro Tyr Leu Ala Arg Ile Thr Glu
 260 265 270
 Thr Arg Glu Leu Phe Ser Ser Lys Glu Arg His Cys Ile His Val Glu
 275 280 285
 Phe Asp Ile Ser Glu Ser Asn Leu Lys Tyr Thr Thr Gly Asp His Leu
 290 295 300
 Ala Ile Trp Pro Ser Asn Ser Asp Glu Asn Ile Lys Gln Phe Ala Lys
 305 310 315 320
 Cys Phe Gly Leu Glu Asp Lys Leu Asp Thr Val Ile Glu Leu Lys Ala
 325 330 335
 Leu Asp Ser Thr Tyr Thr Ile Pro Phe Pro Thr Pro Ile Thr Tyr Gly
 340 345 350
 Ala Val Ile Arg His His Leu Glu Ile Ser Gly Pro Val Ser Arg Gln
 355 360 365
 Phe Phe Leu Ser Ile Ala Gly Phe Ala Pro Asp Glu Glu Thr Lys Lys
 370 375 380
 Thr Phe Thr Arg Leu Gly Gly Asp Lys Gln Glu Phe Ala Thr Lys Val
 385 390 395 400
 Thr Arg Arg Lys Phe Asn Ile Ala Asp Ala Leu Leu Tyr Ser Ser Asn
 405 410 415
 Asn Thr Pro Trp Ser Asp Val Pro Phe Glu Phe Leu Ile Glu Asn Ile
 420 425 430
 Gln His Leu Thr Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Ser Leu Ser
 435 440 445
 Glu Lys Gln Leu Ile Asn Val Thr Ala Val Val Glu Ala Glu Glu Glu
 450 455 460
 Ala Asp Gly Arg Pro Val Thr Gly Val Val Thr Asn Leu Leu Lys Asn
 465 470 475 480
 Ile Glu Ile Ala Gln Asn Lys Thr Gly Glu Lys Pro Leu Val His Tyr
 485 490 495
 Asp Leu Ser Gly Pro Arg Gly Lys Phe Asn Lys Phe Lys Leu Pro Val
 500 505 510
 His Val Arg Arg Ser Asn Phe Lys Leu Pro Lys Asn Ser Thr Thr Pro
 515 520 525
 Val Ile Leu Ile Gly Pro Gly Thr Gly Val Ala Pro Leu Arg Gly Phe
 530 535 540
 Val Arg Glu Arg Val Gln Gln Val Lys Asn Gly Val Asn Val Gly Lys
 545 550 555 560
 Thr Leu Leu Phe Tyr Gly Cys Arg Asn Ser Asn Glu Asp Phe Leu Tyr
 565 570 575
 Lys Gln Glu Trp Ala Glu Tyr Ala Ser Val Leu Gly Glu Asn Phe Glu
 580 585 590
 Met Phe Asn Ala Phe Ser Arg Gln Asp Pro Ser Lys Lys Val Tyr Val
 595 600 605
 Gln Asp Lys Ile Leu Glu Asn Ser Gln Leu Val His Glu Leu Leu Thr
 610 615 620
 Glu Gly Ala Ile Ile Tyr Val Cys Gly Asp Ala Ser Arg Met Ala Arg
 625 630 635 640

Asp Val Gln Thr Thr Ile Ser Lys Ile Val Ala Lys Ser Arg Glu Ile
 645 650 655
 Ser Glu Asp Lys Ala Ala Glu Leu Val Lys Ser Trp Lys Val Gln Asn
 660 665 670
 Arg Tyr Gln Asp Glu Val Trp
 675

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4115 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

CATATGCGCT	AATCTTCTTT	TTCTTTTAT	CACAGGAGAA	ACTATCCCAC	CCCCACTTCG	60
AAACACAATG	ACAACTCCTG	CGTAACTTGC	AAATTCTTGT	CTGACTAATT	GAAAACTCCG	120
GACGAGTCAG	ACCTCCAGTC	AAACGGACAG	ACAGACAAAC	ACTTGGTGCG	ATGTTCATAC	180
CTACAGACAT	GTCAACGGGT	GTTAGACGAC	GGTTTCTTGC	AAAGACAGGT	GTTGGCATCT	240
CGTACGATGG	CAACTGCAGG	AGGTGTCGAC	TTCTCCTTTA	GGCAATAGAA	AAAGACTAAG	300
AGAACAGCGT	TTTACAGGT	TGCAATTGGT	AATGTAGTAT	TTTTTTAGTC	CCAGCATTCT	360
GTGGGTTGCT	CTGGGTTTCT	AGAATAGGAA	ATCACAGGAG	AATGCAAATT	CAGATGGAAG	420
AACAAAGAGA	TAAAAACAA	AAAAAACTG	AGTTTTGCAC	CAATAGAATG	TTTGATGATA	480
TCATCCACTC	GCTAAACGAA	TCATGTGGGT	GATCTTCTCT	TTAGTTTTGG	TCTATCATAA	540
AACACATGAA	AGTGAAATCC	AAATACACTA	CACTCCGGGT	ATTGTCCTTC	GTTTTACAGA	600
TGTCTCATTG	TCTTACTTTT	GAGGTCATAG	GAGTTGCCTG	TGAGAGATCA	CAGAGATTAT	660
CACACTCACA	TTTATCGTAG	TTTCCTATCT	CATGCTGTGT	GTCTCTGGTT	GGTTCATGAG	720
TTTGGATTGT	TGTACATTAA	AGGAATCGCT	GGAAAGCAAA	GCTAACTAAA	TTTTCTTTGT	780
CACAGGTACA	CTAACCTGTA	AAACTTCACT	GCCACGCCAG	TCTTTCCTGA	TTGGGCAAGT	840
GCACAAACTA	CAACCTGCAA	AACAGCACTC	CGCTTGTGAC	AGGTTGTCTC	CTCTCAACCA	900
ACAAAAAAT	AAGATTAAAC	TTTCTTTGCT	CATGCATCAA	TCGGAGTTAT	CTCTGAAAGA	960
GTTGCCTTTG	TGTAATGTGT	GCCAAACTCA	AACTGCAAAA	CTAACACAG	AATGATTTC	1020
CTCACAATTA	TATAAACTCA	CCCACATTTC	CACAGACCGT	AATTTCATGT	CTCACTTTCT	1080
CTTTTGCTCT	TCTTTTACTT	AGTCAGGTTT	GATAACTTCC	TTTTTTATTA	CCCTATCTTA	1140
TTTATTTATT	TATTCATTTA	TACCAACCAA	CCAACCATGG	CCACACAAGA	AATCATCGAT	1200
TCTGTACTTC	CGTACTTGAC	CAAATGGTAC	ACTGTGATTA	CTGCAGCAGT	ATTAGTCTTC	1260
CTTATCTCCA	CAAACATCAA	GAACACGTC	AAGGCAAAGA	AATTGAAATG	TGTCGATCCA	1320
CCATACTTGA	AGGATGCCGG	TCTCACTGGT	ATTCTGTCTT	TGATCGCCGC	CATCAAGGCC	1380
AGAACGACG	GTAGATTGGC	TAACCTTGCC	GATGAAGTTT	TCGACGAGTA	CCCAAACCAC	1440
ACCTTCTACT	TGTCTGTTGC	CGGTGCTTTG	AAGATTGTCA	TGACTGTTGA	CCCAGAAAAC	1500
ATCAAGGCTG	TCTTGCCAC	CCAATTCAC	GACTTCTCCT	TGGGTACCAG	ACACGCCCAC	1560
TTTGCTCCTT	TGTTGGGTGA	CGGTATCTTC	ACCTTGGACG	GAGAAGGTTG	GAAGCACTCC	1620
AGAGCTATGT	TGAGACCACA	GTTTGCTAGA	GACCAGATTG	GACACGTAA	AGCCTTGGAA	1680
CCACACATCC	AAATCATGGC	TAAGCAGATC	AAGTTGAACC	AGGGAAAGAC	TTTCGATATC	1740
CAAGAAATGT	TCTTAGATT	TACCGTCGAC	ACCGCTACTG	AGTTCTTGTT	TGGTGAATCC	1800
GTTCACTCCT	TGTACGATGA	AAAATTGGGC	ATCCCAACTC	CAAACGAAAT	CCCAGGAAGA	1860
GAAAACTTTG	CCGCTGCTTT	CAACGTTTCC	CAACTACT	TGGCCACCAG	AAGTTACTCC	1920
CAGACTTTTT	ACTTTTGGAC	CAACCCTAAG	GAATTCAGAG	ACTGTAACGC	CAAGGTCCAC	1980
CACCTGGCCA	AGTACTTTGT	CAACAAGGCC	TTGAACCTTA	CTCCTGAAGA	ACTCGAAGAG	2040
AAATCCAAGT	CCGGTTACGT	TTTCTGTGAC	GAATTGGTTA	AGCAAACCAG	AGATCCAAAG	2100
GTCTTGCAAG	ATCAATTGTT	GAACATTATG	GTTGCCGGAA	GAGACACCAC	TGCCGGTTTG	2160
TGTCTCTTTG	CTTTGTTTGA	ATTGGCTAGA	CACCCAGAGA	TGTGGTCCAA	GTTGAGAGAA	2220
GAAATCGAAG	TTAACTTTGG	TGTTGGTGAA	GACTCCCGCG	TTGAAGAAAT	TACCTTCGAA	2280
GCCTTGAAGA	GATGTGAATA	CTTGAAGGCT	ATCCTTAACG	AAACCTTGCG	TATGTACCCA	2340
TCGTCTCCTG	TCAACTTTAG	AACCGCCACC	AGAGACACCA	CTTTGCCAAG	AGGTGGTGGT	2400
GCTAACGGTA	CCGACCCAAT	CTACATTCT	AAAGGCTCCA	CTGTTGCTTA	CGTTGTCTAC	2460

AAGACCCACC	GTTTGAAGA	ATACTACGGT	AAGGACGCTA	ACGACTTCAG	ACCAGAAAGA	2520
TGGTTTGAAC	CATCTACTAA	GAAGTTGGGC	TGGGCTTATG	TTCCATTCAA	CGGTGGTCCA	2580
AGAGTCTGCT	TGGGTCAACA	ATTGCGCTTG	ACTGAAGCTT	CTTATGTGAT	CACTAGATTG	2640
GCCCAGATGT	TTGAAACTGT	CTCATCTGAT	CCAGGTCTCG	AATACCCTCC	ACCAAAGTGT	2700
ATTCACITGA	CCATGAGTCA	CAACGATGGT	GTCTTTGTCA	AGATGTAAAG	TAGTCGATGC	2760
TGGGTATTCTG	ATTACATGTG	TATAGGAAGA	TTTTGGTTTT	TTATTCGTTT	TTTTTTTTAA	2820
TTTTTGTAA	ATTAGTTTAG	AGATTTCATT	AATACATAGA	TGGGTGCTAT	TTCCGAAACT	2880
TTACTTCTAT	CCCCTGTATC	CCTTATTATC	CCTCTCAGTC	ACATGATTGC	TGTAATTGTC	2940
GTGAGGACA	CAAACTCCCT	AACGGACTTA	AACCATAAAC	AAGCTCAGAA	CCATAAGCCG	3000
ACATCACTCC	TTCTTCTCTC	TTCTCCAACC	AATAGCATGG	ACAGACCCAC	CCTCCTATCC	3060
GAATCGAAGA	CCCTTATTGA	CTCCATACCC	ACCTGGAAGC	CCCTCAAGCC	ACACACGTCA	3120
TCCAGCCAC	CCATCACCAC	ATCCCTCTAC	TGACAAACGT	CCAAAGACGG	CGAGTTCTGG	3180
TGTGCCCCGA	AATCAGCCAT	CCCGGCCACA	TACAAGCAGC	CGTTGATTGC	GTGCATACTC	3240
GGCGAGCCCA	CAATGGGAGC	CACGCATTCTG	GACCATGAAG	CAAAGTACAT	TCACGAGATC	3300
ACGGGTGTTT	CAGTGTGCGA	GATTGAGAAG	TTCGACGATG	GATGGAAGTA	CGATCTCGTT	3360
GCGGATTACG	ACTTCGGTGG	GTGTATTCT	AAACGAAGAT	TCTATGAGAC	GCAGCATGTG	3420
TTTCGGTTCG	AGGATTGTGC	GTACGTCATG	AGTGTGCCCT	TTGATGGACC	CAAGGAGGAA	3480
GGTTACGTGG	TTGGGACGTA	CAGATCCATT	GAAAGTTGA	GCTGGGGTAA	AGACGGGGAC	3540
GTGGAGTGGG	CCATGGCGAC	GACGTCGGAT	CCTGGTGGGT	TTATCCCGCA	ATGGATAACT	3600
CGATTGAGCA	TCCTTGAGC	AATCGCAAAA	GATGTGCCCTA	GTGTATTAA	CTACATACAG	3660
AAATAAAAC	GTGTCTTGAT	TCATTGGTTT	GGTCTTGT	GGGTCCGAG	CCAATATTTT	3720
ACATCATCTC	CTAAATCTC	CAAGAATCCC	AACGTAGCGT	AGTCCAGCAC	GCCCTCTGAG	3780
ATCTTATTTA	ATATCGACTT	CTCAACCACC	GGTGAATCC	CGTTCAGACC	ATTGTTACCT	3840
GTAGTGTGTT	TGCTCTTGTT	CTTGATGACA	ATGATGTATT	TGTCACGATA	CCTGAAATAA	3900
TAAACATCC	AGTCATTGAG	CTTATTACTC	GTGAACCTAT	GAAAGAACTC	ATTCAGCCCG	3960
TTCCCAAAA	ACCCAGAATT	GAAGATCTTG	CTCAACTGGT	CATGCAAGTA	GATGATCGCC	4020
ATGATCTGAT	ACTTTACCAA	GCTATCTCT	CCAAGTTCTC	CCACGTACGG	CAAGTACGGC	4080
AACGAGCTCT	GGAAGCTTTG	TTGTTGGGG	TCATA			4115

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3948 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

GACCTGTGAC	GCTTCCGGTG	TCTTGCCACC	AGTCTCCAAG	TTGACCGACG	CCCAAGTCAT	60
GTACCACITT	ATTTCGGTT	ACACTTCCAA	GATGGCTGGT	ACTGAAGAAG	GTGTCACGGA	120
ACCACAAGCT	ACTTCTCCG	CTTGTTTCGG	TCAACCATT	TTGGTGTGTC	ACCCAAATGAA	180
GTACGCTCAA	CAATTGTCTG	ACAAGATCTC	GCAACACAAG	GCTAACGCCT	GGTTGTTGAA	240
CACCGTTGG	GTGGTTCTT	CTGCTGCTAG	AGGTGGTAAG	AGATGCTCAT	TGAAGTACAC	300
CAGAGCCATT	TTGGACGCTA	TCCACTCTGG	TGAATTGTCC	AAGGTTGAAT	ACGAACTTT	360
CCCAGTCTTC	AACTTGAATG	TCCCAACCTC	CTGTCCAGGT	GTCCCAAGTG	AAATCTTGAA	420
CCCAACCAAG	GCCTGGACCG	GAAGGTGTTG	ACTCCTTCAA	CAAGGAAATC	AAGTCTTTGG	480
CTGGTAAGTT	TGCTGAAAAC	TTCAAGACCT	ATGCTGACCA	AGCTACCGCT	GAAGTGAGAG	540
CTGCAGGTCC	AGAAGCTTAA	AGATATTTAT	TCATTATTTA	GTTTGCCTAT	TTATTTCTCA	600
TTACCCATCA	TCATTCAACA	CTATATATAA	AGTTACTTCG	GATATCATTG	TAATCGTGCG	660
TGTGCAATT	GGATGATTG	GAACTGCGCT	TGAAACGGAT	TCATGCACGA	AGCGGAGATA	720
AAAGATTACG	TAATTTATCT	CCTGAGACAA	TTTTAGCCGT	GTTTACACGC	CCTTCTTTGT	780
TCTGAGCGAA	GGATAAATAA	TTAGACTTCC	ACAGCTCATT	CTAATTTCCG	TCACGCGAAT	840
ATTGAAGGGG	GGTACATGTG	GCCGCTGAAT	GTGGGGGCAG	TAAACGCAGT	CTCTCCTCTC	900
CCAGGAATAG	TGCAACGGAG	GAAGGATAAC	GGATAGAAAG	CGGAATGCGA	GGAAAAATTT	960
GAACGCGCAA	GAAAAGCAAT	ATCCGGGCTA	CCAGGTTTTG	AGCCAGGGAA	CACACTCCTA	1020
TTTCTGCTCA	ATGACTGAAC	ATAGAAAAAA	CACCAAGACG	CAATGAAACG	CACATGGACA	1080
TTTAGACCTC	CCCACATGTG	ATAGTTTGTG	TTAACAGAAA	AGTATAATAA	GAACCCATGC	1140

CGTCCCTTTT	CTTTCGCCGC	TTCAACTTTT	TTTTTTTTAT	CTTACACACA	TCACGACCAT	1200
GACTGTACAC	GATATTATCG	CCACATACTT	CACCAAATGG	TACGTGATAG	TACCACTCGC	1260
TTTGATTGCT	TATAGAGTCC	TCGACTACTT	CTATGGCAGA	TACTTGATGT	ACAAGCTTGG	1320
TGCTAAACCA	TTTTTCCAGA	AACAGACAGA	CGGCTGTTTC	GGATTCAAAG	CTCCGCTTGA	1380
ATTGTTGAAG	AAGAAGAGCG	ACGGTACCCT	CATAGACTTC	ACACTCCAGC	GTATCCACGA	1440
TCTCGATCGT	CCCATATATCC	CAACTTTTAC	ATTCGCCGTC	TTTTCCATCA	ACCTTGTCAA	1500
TACCCCTGAG	CCGGAGAACA	TCAAGGCCAT	CTTGGCCACT	CAGTTCAACG	ATTTCTCCTT	1560
GGGTACCAGA	CACCTCGACT	TTGCTCCTTT	GTTGGGTGAT	GGTATCTTTA	CGTTGGATGG	1620
CGCCGGCTGG	AAGCACAGCA	GATCTATGTT	GAGACCACAG	TTTGCCAGAG	AACAGATTTT	1680
CCACGTCAAG	TTGTTGGAGC	CACACGTTCA	GGTGTCTTTC	AAACACGTCA	GAAAGGCACA	1740
GGGCAAGACT	TTTGACATCC	AGGAATTGTT	TTTCAGATTG	ACCGTCGACT	CCGCCACCGA	1800
GTTTTTGT	GGTGAATCCG	TTGAGTCCTT	GAGAGATGAA	TCTATCGGCA	TGTCCATCAA	1860
TGCGCTTGAC	TTTGACGGCA	AGGCTGGCTT	TGCTGATGCT	TTTAACTATT	CGCAGAATTA	1920
TTTGGCTTCG	AGAGCGGTTA	TGCAACAATT	GTAAGGCTG	TTGAACGGGA	AAAAGTTTAA	1980
GGAGTGCAAC	GCTAAAGTGC	ACAAGTTTGC	TGACTACTAC	GTCACCAAGG	CTTTGGACTT	2040
GACGCTGAA	CAATTGGA	AGCAGGATGG	TTATGTGTTT	TTGTACGAAT	TGGTCAAGCA	2100
AACCAGAGAC	AAGCAAGTGT	TGAGAGACCA	ATTGTTGAAC	ATCATGGTTG	CTGGTAGAGA	2160
CACCACCGCC	GGTTTGTGTT	CGTTTGTGTT	CTTTGAATTG	GCCAGAAACC	CAGAAGTTAC	2220
CAACAAGTTG	AGAGAAGAAA	TTGAGGACAA	GTTTGGACTC	GGTGAGAATG	CTAGTGTGTA	2280
AGACATTTCC	TTTGAGTCGT	TGAAGTCCTG	TGAATACTTG	AAGGCTGTTC	TCAACGAAAC	2340
CTTGAGATTG	TACCCATCCG	TGCCACAGAA	TTTCAGAGTT	GCCACCAAGA	ACACTACCCT	2400
CCCAAGAGGT	GGTGGTAAGG	ACGGGTGTC	TCCTGTTTGG	GTGAGAAAGG	GTCAGACCGT	2460
TATTTACGGT	GTCTACGCAG	CCCACAGAAA	CCCAGCTGTT	TACGGTAAGG	ACGCTCTTGA	2520
GTTTAGACCA	GAGAGATGGT	TTGAGCCAGA	GACAAAGAAG	CTTGGCTGGG	CCTTCTCTCC	2580
ATTCAACGGT	GGTCCAAGAA	TCTGTTTGGG	ACAGCAGTTT	GCCTTGACAG	AAGCTTCGTA	2640
TGTCAGTCTC	AGGTTGCTCC	AGGAGTTTGC	ACACTTGTCT	ATGGACCCAG	ACACCGAATA	2700
TCCACCTAAG	AAAATGTGCG	ATTTGACCAT	GTGCTTTTTC	GACGGTGCCA	ATATTGAGAT	2760
GTATTAGAGG	GTGATGTGTT	ATTTTGATTG	TTTAGTTTGT	AATTACTGAT	TAGGTTAATT	2820
CATGGATTGT	TATTTATTGA	TAGGGGTTTG	CGCGTGTGTC	ATTCACCTGG	GATCGTTCCA	2880
GGTTGATGTT	TCCTTCCATC	CTGTGAGTGC	AAAAGGAGTT	TTGTTTGTGA	ACTCCGGACG	2940
ATGTTTTTAA	TAGAAGGTCG	ATCTCCATGT	GATTGTTTTG	ACTGTTACTG	TGATTATGTA	3000
ATCTGCGGAC	GTTATACAAG	CATGTGATTG	TGGTTTTGCA	GCCTTTTGCA	CGACAAATGA	3060
TCGTGAGACG	ATTACGTAAT	CTTTGTTAGA	GGGGTAAAAA	AAAACAAAT	GGCAGCCAGA	3120
ATTTCAAACA	TTCTGCAAAC	AATGCAAAAA	ATGGGAAACT	CCAACAGACA	AAAAAAAAAA	3180
CTCCGCAGCA	CTCCGAACCC	ACAGAACAA	GGGGCGCCAG	AATTATTGAC	TATTGTGACT	3240
TTTTTACGCT	AACGCTCATT	GCAGTGTAGT	GCGTCTTACA	CGGGGTATTG	CTTTCTACAA	3300
TGCAAGGGCA	CAGTTGAAGG	TTTGACCTA	ACGTTGCCCC	GTGTCAACTC	AATTTGACGA	3360
GTAACCTTCT	AAGCTCGAAT	TATGCAGCTC	GTGCGTCAAC	CTATGTGCAG	GAAAGAAAAA	3420
ATCCAAAAA	ATCGAAAAATG	CGACTTTCGA	TTTTGAATAA	ACCAAAAAAG	AAAATGTGCG	3480
ACTTTTTTCT	CGCTCTCGCT	CTCTCGACCC	AAATCACAAC	AAATCCTCGC	GCGCAGTATT	3540
TCGACGAAAC	CACAACAAAT	AAAAAAACA	AATTCTACAC	CACTTCTTTT	TCTTCACCAG	3600
TCAACAAAAA	ACAACAAATT	ATACACCATT	TCAACGATTT	TTGCTCTTAT	AAATGCTATA	3660
TAATGGTTTA	ATTCAACTCA	GGTATGTTTA	TTTTACTGTT	TTCAGCTCAA	GTATGTTCAA	3720
ATACTAATA	CTTTTGATGT	TTGTCGCTTT	TCTAGAATCA	AAACAACGCC	CACAACACGC	3780
CGAGCTTGTC	GAATAGACGG	TTTGTGTTACT	CATTAGATGG	TCCCAGATTA	CTTTTCAAGC	3840
CAAAGTCTCT	CGAGTTTTGT	TTGCTGTTTC	CCCAATTCCT	AACTATGAAG	GGTTTTTATA	3900
AGGTCCAAAG	ACCCCAAGGC	ATAGTTTTTT	TGGTTCCTTC	TTGTCGTG		3948

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3755 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GCTCAACAAT	TGCTGACAA	GATCTCGCAA	CACAAGGCTA	ACGCCTGGTT	GTTGAACACT	60
GGTTGGGTTG	GTTCTTCTGC	TGCTAGAGGT	GGTAAGAGAT	GTTCAATTGAA	GTACACCAGA	120
GCCATTTTGG	ACGCTATCCA	CTCTGGTGAA	TTGTCCAAGG	TTGAATACGA	GACTTTCCCA	180
GTCTTCAACT	TGAATGTCCC	AACCTCCTGC	CCAGGTGTCC	CAAGTGAAAT	CTTGAACCCA	240
ACCAAGGCCT	GGACCGAAGG	TGTTGACTCC	TTCAACAAGG	AAATCAAGTC	TTGGGCTGGT	300
AAGTTTGCTG	AAAACCTTCAA	GACCTATGCT	GACCAAGCTA	CCGCTGAAGT	TAGAGCTGCA	360
GGTCCAGAAG	CTTAAAGATA	TTTATTCACT	ATTTAGTTTG	CCTATTTATT	TCTCATCACC	420
CATCATCATT	CAACAATATA	TATAAAGTTA	TTTCGGAAC	CATATATCAT	TGTAATCGTG	480
CGTGTGCAA	TTGGGTAATT	TGAACTGTA	GTTGGAACGG	ATTCAATGCAC	GATGCGGAGA	540
TAACACGAGA	TTATCTCCTA	AGACAATTTT	GGCCTCATTC	ACACGCCCTT	CTTCTGAGCT	600
AAGGATAAAT	AATTAGACTT	CACAAGTTCA	TTAAAATATC	CGTCACGCGA	AAACTGCAAC	660
AATAAGGAAG	GGGGGGGTAG	ACGTAGCCGA	TGAATGTGGG	GTGCCAGTAA	ACGCAGTCTC	720
TCTCTCCCCC	CCCCCCCCC	CCCCCTCAGG	AATAGTACAA	CGGGGGAAGG	ATAACGGATA	780
GCAAGTGAAG	TGCGAGGAAA	ATTTTGAATG	CGCAAGGAAA	GCAATATCCG	GGCTATCAGG	840
TTTTGAGCCA	GGGGACACAC	TCCTCTTCTG	CACAAAAACT	TAACGTAGAC	AAAAAAAAAA	900
AACTCCACCA	AGACACAATG	AATCGCACAT	GGACATTTAG	ACCTCCCCAC	ATGTGAAAGC	960
TTCTCTGGCG	AAAGCAAAAA	AAGTATAATA	AGGACCCATG	CCTTCCCTCT	TCCTGGGCCG	1020
TTTCAACTTT	TTCTTTTTCT	TTGTCTATCA	ACACACACAC	ACCTCACGAC	CATGACTGCA	1080
CAGGATATTA	TCGCCACATA	CATCACCAAA	TGGTACGTGA	TAGTACCACT	CGCTTTGATT	1140
GCTTATAGGG	TCCTCGACTA	CTTTTACGGC	AGATACTTGA	TGTACAAGCT	TGGTGTCTAAA	1200
CCGTTTTTCC	AGAAACAAAC	AGACGGTTAT	TTCCGATTCA	AAGCTCCACT	TGAATTGTTA	1260
AAAAAGAAGA	GTGACGGTAC	CCTCATAGAC	TTCACTCTCG	AGCGTATCCA	AGCGCTCAAT	1320
CGTCCAGATA	TCCCAACTTT	TACATTCCCA	ATCTTTTCCA	TCAACCTTAT	CAGCACCCCT	1380
GAGCCGAGAG	ACATCAAGGC	TATCTTGGCC	ACCCAGTTCA	ACGATTTCTC	CTTGGGCACC	1440
AGACACTCGC	ACTTTGCTCC	TTTGTGGGCC	GATGGTATCT	TTACCTTGGA	CGGTGCCGGC	1500
TGGAAGCACA	GCAGATCTAT	GTTGAGACCA	CAGTTTGCCA	GAGAACAGAT	TTCCCACGTC	1560
AAGTTGTGGG	AGCCACACAT	GCAGGTGTTT	TTCAAGCACG	TCAGAAAGGC	ACAGGGCAAG	1620
ACTTTTGACA	TCCAAGAATT	GTTTTTCAGA	TTGACCGTCG	ACTCCGCCAC	TGAGTTTTTG	1680
TTTGGTGAAT	CCGTTGAGTC	CTTGAGAGAT	GAATCTATTG	GGATGTCCAT	CAATGCACTT	1740
GACTTTGACG	GCAAGGCTGG	CTTTGCTGAT	GCTTTTAACT	ACTCGCAGAA	CTATTTGGCT	1800
TCGAGAGCGG	TTATGCAACA	ATTGTAAGTG	GTGTTGAACG	GGAAAAAGTT	TAAGGAGTGC	1860
AACGCTAAAG	TGCACAAGTT	TGCTGACTAT	TACGTCAGCA	AGGCTTTGGA	CTTGACACCT	1920
GAACAATTGG	AAAAGCAGGA	TGGTTATGTG	TTCTTGTAAG	AGTTGGTCAA	GCAAACCAGA	1980
GACAGGCAAG	TGTTGAGAGA	CCAGTTGTTG	AACATCATGG	TTGCCGGTAG	AGACACCACC	2040
GCCGGTTTGT	TGTCGTTTGT	TTTCTTTGAA	TTGGCCAGAA	ACCCAGAGGT	GACCAACAAG	2100
TTGAGAGAAG	AAATCGAGGA	CAAGTTTGGT	CTTGGTGAGA	ATGCTCGTGT	TGAAGACATT	2160
TCCTTTGAGT	CGTTGAAGTC	ATGTGAATAC	TTGAAGGCTG	TTCTCAACGA	AACTTTGAGA	2220
TTGTACCCAT	CCGTGCCACA	GAATTTGAGA	GTTGCCACCA	AAAACACTAC	CCTTCCAAGG	2280
GGAGGTGGTA	AGGACGGGTT	ATCTCCTGTT	TTGGTCAGAA	AGGGTCAAAC	CGTTATGTAC	2340
GGTGTCTACG	CTGCCCCACG	AAACCCAGCT	GTCTACGGTA	AGGACGCCCT	TGAGTTTAGA	2400
CCAGAGAGGT	GGTTTGAGCC	AGAGACAAAG	AAGCTTGGCT	GGGCCCTTCT	TCCATTCAAC	2460
GGTGGTCCAA	GAATTTGCTT	GGGACAGCAG	TTTGCCTTGA	CAGAAGCTTC	GTATGTCACT	2520
GTCAGATTGC	TCCAAGAGTT	TGGACACTTG	TCTATGGACC	CCAACACCGA	ATATCCACCT	2580
AGGAAAAATG	CGCATTTGAC	CATGTCCCTT	TTCCAGCGTG	CCAACATTGA	GATGTATTAG	2640
AGGATCATGT	GTTATTTTTG	ATTGGTTTAG	TCTGTTTGTA	GCTATTGATT	AGGTTAATTC	2700
ACGGATTGTT	ATTTATTGAT	AGGGGGTGCG	TGTGTGTGTG	TGTGTTGCAT	TCACATGGGA	2760
TCGTTCCAGG	TTGTTGTTTC	CTTCCATCCT	GTTGAGTCAA	AAGGAGTTTT	GTTTTGTAAC	2820
TCCGGACGAT	GTCTTAGATA	GAAGGTCGAT	CTCCATGTGA	TTGTTTGACT	GCTACTCTGA	2880
TTATGTAATC	TGTAAAGCCT	AGACGTTATG	CAAGCATGTG	ATTGTGGTTT	TTGCAACCTG	2940
TTTGACGAC	AAATGATCGA	CAGTCGATTA	CGTAATCCAT	ATTATTTAGA	GGGGTAATAA	3000
AAAATAAATG	GCAGCCAGAA	TTTCAAACAT	TTTGCAAACA	ATGCAAAAGA	TGAGAAACTC	3060
CAACAGAAAA	AATAAAAAAA	CTCCGCAGCA	CTCCGAACCA	ACAAAAAAT	GGGGGGCGCC	3120
AGAATTATTG	ACTATTGTGA	CTTTTTTTTA	TTTTTCCGT	TAACTTTCAT	TGCAGTGAAG	3180
TGTGTTACAC	GGGGTGGTGA	TGGTGTGGT	TTCTACAATG	CAAGGGCACA	GTTGAAGGTT	3240
TCCACATAAC	GTTGCACCAT	ATCRACTCAA	TTTATCCTCA	TTCATGTGAT	AAAAGAAGAG	3300

CCAAAAGGTA	ATTGGCAGAC	CCCCAAGGG	GAACACGGAG	TAGAAAGCAA	TGGAAACACG	3360
CCCATGACAG	TGCCATTTAG	CCCACAACAC	ATCTAGTATT	CTTTTTTTTT	TTTGTGCGCA	3420
GGTGCACACC	TGGACTTTAG	TTATTGCCCC	ATAAAGTTAA	CAATCTCACC	TTTGGCTCTC	3480
CCAGTGTCTC	CGCCTCCAGA	TGCTCGTTTT	ACACCCTCGA	GCTAACGACA	ACACAACACC	3540
CATGAGGGGA	ATGGGCAAAG	TTAAACACTT	TTGGTTTCAA	TGATTCCTAT	TTGCTACTCT	3600
CTTGTTTTGT	GTTTTGATTT	GCACCATGTG	AAATAAACGA	CAATTATATA	TACCTTTTCG	3660
TCTGTCTCC	AATGTCTCTT	TTTGCTGCCA	TTTGCTTTT	TGCTTTTTGC	TTTTGCACTC	3720
TCTCCCACTC	CCACAATCAG	TGCAGCAACA	CACAA			3755

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3900 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GACATCATAA	TGACCCGGTT	ATTTGCGCCT	CAGGTTGCTT	ATTTGAGCCG	TAAAGTGCAG	60
TAGAACTTT	GCCTTGGGTT	CAAACCTAG	TATAATGGTG	ATAACTGGTT	GCACTCTTGC	120
CATAGGCATG	AAAATAGGCC	GTTATAGTAC	TATATTTAAT	AAGCGTAGGA	GTATAGGATG	180
CATATGACCG	GTTTTTCTAT	ATTTTTAAGA	TAATCTCTAG	TAAATTTTGT	ATTCTCAGTA	240
GGATTTTCATC	AAATTTTCGCA	ACCAATTCTG	GCGAAAAAAT	GATTCTTTTA	CGTCAAAGC	300
TGAATAGTGC	AGTTTAAAGC	ACCTAAATC	ACATATACAG	CCTCTAGATA	CGACAGAGAA	360
GCTCTTTATG	ATCTGAAGAA	GCATTAGAAT	AGCTACTATG	AGCCACTATT	GGTGTATATA	420
TTAGGGATTG	GTGCAATTAA	GTACGTACTA	ATAAACAGAA	GAAAATACTT	AACCAATTTT	480
TGGTGTATAC	TTAGTGGTGA	GGGACCTTTT	CTGAACATTC	GGGTCAAAC	TTTTTTTGGA	540
GTGCGACATC	GATTTTTTCGT	TGTGTAAATA	ATAGTGAACC	TTTGTGTAAT	AAATCTTCAT	600
GCAAGACTTG	CATAATTCTGA	GCTTGGGAGT	TCACGCCAAT	TTGACCTCGT	TCATGTGATA	660
AAAGAAAAGC	CAAAGGTAA	TTAGCAGACG	CAATGGGAAC	ATGGAGTGGG	AAGCAATGGA	720
AGCACGCCCA	GGACGGAGTA	ATTTAGTCCA	CACTACATCT	GGGGGTTTTT	TTTTTGTGCG	780
CAAGTACACA	CCTGGACTTT	AGTTTTTGCC	CCATAAAGTT	AACAATCTAA	CCTTTGGCTC	840
TCCAACTCTC	TCCGCCCCCA	AATATTCGTT	TTTACACCCT	CAAGCTAGCG	ACAGCACAAC	900
ACCCATTAGA	GGAATGGGGC	AAAGTTAAAC	ACTTTTGGCT	TCAATGATTC	CTATTCGCTA	960
CTACATTCCT	CTCTTGTTTT	GTGCTTGAA	TTGCACCATG	TGAAATAAAC	GACAATTATA	1020
TATACCTTTT	CATCCCTCCT	CCTATATCTC	TTTTTGCTAC	ATTTTGTTTT	TTACGTTTCT	1080
TGCTTTTGCA	CTCTCCCACT	CCCACAAAGA	AAAAAAACT	ACACTATGTC	GTCTTCTCCA	1140
TCGTTTGCCC	AAGAGGTTCT	CGCTACCACT	AGTCCTTACA	TCGAGTACTT	TCTTGACAAC	1200
TACACCAGAT	GGTACTACTT	CATACCTTGT	GTGCTTCTTT	CGTTGAACCT	TATAAGTTTG	1260
CTCCACACAA	GGTACTTGGA	ACGCAGGTTT	CACGCCAAGC	CACTCGGTAA	CTTTGTCAGG	1320
GACCCTACGT	TTGGTATCGC	TACTCCGTTG	CTTTTGATCT	ACTTGAAGTC	GAAAGGTACG	1380
GTCTGAAGT	TTGCTTGGGG	CCTCTGGAAC	AACAAGTACA	TCGTGAGAGA	CCCAAAGTAC	1440
AAGACAACCTG	GGCTCAGGAT	TGTTGGCCTC	CCATTGATTG	AAACCATGGA	CCCAGAGAAC	1500
ATCAAGGCTG	TTTTGGCTAC	TCAGTTCAAT	GATTTCTCTT	TGGGAACCAG	ACACGATTTT	1560
TTGTACTCCT	TGTTGGGTGA	CGGTATTTTC	ACCTTGGACG	GTGCTGGCTG	GAAACATAGT	1620
AGAACTATGT	TGAGACCACA	GTTTGCTAGA	GAACAGGTTT	CTCACGTCAA	GTTGTGGGAG	1680
CCACACGTTT	AGGTGTTCTT	CAAGCACGTT	AGAAAGCACC	GCGGTCAAAC	GTTTCGACATC	1740
CAAGAATTGT	TCTTCAGGTT	GACCGTCGAC	TCCGCCACCG	AGTTCTTGTT	TGGTGAGTCT	1800
GCTGAATCCT	TGAGGGACGA	ATCTATTGGA	TTGACCCCAA	CCACCAAGGA	TTTCGATGGC	1860
AGAAGAGATT	TCGCTGACGC	TTTCAACTAT	TCGCAGACTT	ACCAGGCCTA	CAGATTTTTG	1920
TTGCAACAAA	TGTACTGGAT	CTTGAATGGC	TCGGAATTCA	GAAAGTCGAT	TGCTGTCGTG	1980
CACAAGTTTG	CTGACCACTA	TGTGCAAAAG	GCTTTGGAGT	TGACCGACGA	TGACTTGCAG	2040
AAACAAGACG	GCTATGTGTT	CTTGACGAG	TTGGCTAAGC	AAACCAGAGA	CCCAAAGGTC	2100
TTGAGAGACC	AGTTATTGAA	CATTTTGGTT	GCCGGTAGAG	ACACGACCGC	CGGTTTGTG	2160
TCATTTGTTT	TCTACGAGTT	GTCAAGAAAC	CCTGAGGTGT	TTGCTAAGTT	GAGAGAGGAG	2220
TGTGAAAACA	GATTTGGACT	CGGTGAAGAA	GCTCGTGTG	AAGAGATCTC	GTTTGAGTCC	2280
TTGAAGTCTT	GTGAGTACTT	GAAGGCTGTC	ATCAATGAAA	CCTTGAGATT	GTACCCATCG	2340

GTTCCACACA	ACTTTAGAGT	TGCTACCAGA	AACACTACCC	TCCCAAGAGG	TGGTGGTGAA	2400
GATGGATACT	CGCCAATTGT	CGTCAAGAAG	GGTCAAGTTG	TCATGTACAC	TGTTATTGCT	2460
ACCCACAGAG	ACCCAAGTAT	CTACGGTGCC	GACGCTGACG	TCTTCAGACC	AGAAAGATGG	2520
TTTGAAACCAG	AAACTAGAAA	GTTGGGCTGG	GCATACGTTT	CATTCAATGG	TGGTCCAAGA	2580
ATCTGTTTGG	GTCAACAGTT	TGCCTTGACC	GAAGCTTCAT	ACGTCACGTG	CAGATTGCTC	2640
CAGGAGTTTG	CACACTTGTC	TATGGACCCA	GACACCGAAT	ATCCACCAAA	ATTGCAGAAC	2700
ACCTTGACCT	TGTCGCTCTT	TGATGGTGCT	GATGTTAGAA	TGTACTAAGG	TTGCTTTTCC	2760
TTGCTAAATT	TCTTCTGTAT	AGCTTGTGTA	TTTAAATTGA	ATCGGCAATT	GATTTTTCTG	2820
ATACCAATAA	CCGTAGTGCG	ATTTGACCAA	AACCGTTCAA	ACTTTTTGTT	CTCTCGTTGA	2880
CGTGCTCGCT	CATCAGCACT	GTTTGAAGAC	GAAAGAGAAA	ATTTTTTGTA	AACAACACTG	2940
TCCAAATTTA	CCCAACGTGA	ACCATTATGC	AAATGAGCGG	CCCTTTC AAC	TGGTCGCTGG	3000
AAGCATTGCG	GGATATCTAC	AACGCCCTTA	AGTTTGAAAC	AGACATTGAT	TTAGACACCA	3060
TAGATTTTCA	CGGCATCAAG	AATGACCTTG	CCCACATTTT	GACGACCCCA	ACACCACTGG	3120
AAGAATCAGC	CCAGAACTA	GGCGATGGAT	CCAAGCCTGT	GACCTTGCCC	AATGGAGACG	3180
AAGTGGAGTT	GAACCAAGCG	TTCCTAGAAG	TTACCACATT	ATTGTCGAAT	GAGTTTGACT	3240
TGGACCAATT	GAACGCGGCA	GAGTTGTTAT	ACTACGCTGG	CGACATATCC	TACAAGAAGG	3300
GCACATCAAT	CGCAGACAGT	GCCAGATTGT	CTTATTATTT	GAGAGCAAAC	TACATCTTGA	3360
ACATACTTGG	GTATTTGATT	TCGAAGCAGC	GATTGGATTT	GATAGTCACG	GACAACGACG	3420
CGTTGTTTGA	TAGTATTTTG	AAAAGTTTGG	AAAAGATCTA	CAAGTTGATA	AGCGTGTTGA	3480
ACGATATGAT	TGACAAGCAA	AAGGTGACAA	GCGACATCAA	CAGTCTAGCA	TTCATCAATT	3540
GCATCAACTA	CTCGAGAGGT	CAACTATTCT	CCGCACACGA	ACTTTTGGGA	CTGGTTTGTG	3600
TTGGATTGGT	CGACATCTAT	TTCACCCAGT	TTGGCACATT	AGACAACACT	AAGAAGGTAT	3660
TGGCATTGAT	ACTGAAGAAC	ATCAGCGATG	AAGACATCTT	GATCATACAC	TTCTCCCAT	3720
CGACACTACA	ATTGTTTAAG	CTGGTGTGG	ACAAGAAAGA	CGACGCTGCA	GTGAACAGT	3780
TCTACAAGTA	CATCACTTCA	ACAGTGTAC	GAGACTACAA	CTCCAACATC	GGCTCCACAG	3840
CCAAAGATGA	TATCGATTG	TCCAAAACCA	AACCTCAGTG	CTTTGAGGTG	TTGACGAGTT	3900

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3668 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

CCTGCAGAAT	TCGCGGCCGC	GTGACAGAG	TAGCAGTTAT	GCAAGCATGT	GATTGTGGTT	60
TTTGCAACCT	GTTTGCACGA	CAAATGATCG	ACAGTCGATT	ACGTAATCCA	TATTATTTAG	120
AGGGGTAAAT	AAAAATAAAT	GGCAGCCAGA	ATTTCAAACA	TTTTGC AAAAC	AATGCAAAAG	180
ATGAGAAACT	CCACAGAAA	AAATAAAAAA	ACTCCGCAGC	ACTCCGAACC	AACAAAACAA	240
TGGGGGGCGC	CAGAATTATT	GACTATTGTG	ACTTTTTTTT	ATTTTTTCCG	TTAACTTTCA	300
TTGCAGTGAA	GTGTGTTACA	CGGGGTGGTG	ATGGTGTGG	TTTCTACAAT	GCAAGGGCAC	360
AGTTGAAGGT	TTCCACATAA	CGTTGCACCA	TATCAACTCA	ATTTATCCTC	ATTCATGTGA	420
TAAAAGAAGA	GCCAAAAGGT	AATTGGCAGA	CCCCCAAGG	GGAACACGGA	GTAAGAAAGCA	480
ATGGAACAC	GCCCATGACA	GTGCCATTTA	GCCCAACA	CATCTAGTAT	TCITTTTTTT	540
TTTGTGCGC	AGGTGCACAC	CTGGACTTTA	GTTATTGCCC	CATAAAGTTA	ACAATCTCAC	600
CTTTGGCTCT	CCAGTGTCT	CCGCCTCCAG	ATGCTCGTTT	TACACCCTCG	AGCTAACGAC	660
AACACAACAC	CCATGAGGGG	AATGGGCAAA	GTAAACACT	TTTGGTTTCA	ATGATTCCCTA	720
TTTGCTACTC	TCITGTTTTG	TGTTTGTATT	TGCACCATGT	GAAATAAAG	ACAATTATAT	780
ATACCTTTTC	GTCTGTCTCT	CAATGTCTCT	TTTTGCTGCC	ATTTTGCTTT	TTGCTTTTTG	840
CTTTTGCAC	CTCTCCCACT	CCCACAATCA	GTGCAGCAAC	ACACAAAGAA	GAAAAATAAA	900
AAAACCTACA	CTATGTCGTC	TTCTCCATCG	TTTGCTCAGG	AGGTTCTCGC	TACCACTAGT	960
CCTTACATCG	AGTACTTTCT	TGACAACACT	ACCAGATGGT	ACTACTTCAT	CCCTTTGGTG	1020
CTTCTTTCTG	TGAACCTCAT	CAGCTTGCTC	CACACAAAGT	ACTTGGAAAG	CAGGTTCCAC	1080
GCCAAGCCGC	TCGGTAACGT	CGTGTGGAT	CCTACGTTTG	GTATCGCTAC	TCCGTTGATC	1140
TTGATCTACT	TAAAGTCGAA	AGGTACAGTC	ATGAAGTTTG	CCTGGAGCTT	CTGGAACAAC	1200
AAGTACATTG	TCAAAGACCC	AAAGTACAAG	ACCCTGGGCC	TTAGAATTGT	CGGCCTCCCA	1260

TTGATTGAAA	CCATAGACCC	AGAGAACATC	AAAGCTGTGT	TGGCTACTCA	GTTCAACGAT	1320
TTCTCCTTGG	GAAGTAGACA	CGATTTCTTG	TACTCCTTGT	TGGGCGATGG	TATTTTACC	1380
TTGGACGGTG	CTGGCTGGAA	ACACAGTAGA	ACTATGTTGA	GACCACAGTT	TGCTAGAGAA	1440
CAGGTTTCCC	ACGTCAAGTT	GTTGGAACCA	CACGTTCCAG	TGTTCTTCAA	GCACGTTAGA	1500
AAACACCGCG	GTCAGACTTT	TGACATCCAA	GAATTGTTCT	TCAGATTGAC	CGTCGACTCC	1560
GCCACCGAGT	TCTTGTGTTG	TGAGTCTGCT	GAATCCTTGA	GAGACGACTC	TGTTGGTTTG	1620
ACCCCAACCA	CCAAGGATTT	CGAAGGCAGA	GGAGATTTCG	CTGACGCTTT	CAACTACTCG	1680
CAGACTTACC	AGGCCTACAG	ATTTTTGTTG	CAACAAATGT	ACTGGATTTT	GAATGGCGCG	1740
GAATTCAGAA	AGTCGATTGC	CATCGTGCAC	AAGTTTGCTG	ACCACTATGT	GCAAAAGGCT	1800
TTGGAGTTGA	CCGACGATGA	CTTGACAGAA	CAAGACGGCT	ATGTGTTCTT	GTACGAGTTG	1860
GCTAAGCAAA	CTAGAGACCC	AAAGGTCTTG	AGAGACCAGT	TGTTGAACAT	TTTGGTTGCC	1920
GGTAGAGACA	CGACCGCCGG	TTTGTGTCG	TTTGTGTTCT	ACGAGTTGTC	GAGAAACCCCT	1980
GAAGTGTTTG	CCAAGTTGAG	AGAGGAGGTG	GAAAACAGAT	TTGGACTCGG	CGAAGAGGCT	2040
CGTGTTGAAG	AGATCTCTTT	TGAGTCCTTG	AAGTCCTGTG	AGTACTTGAA	GGCTGTCTATC	2100
AATGAAGCCT	TGAGATTGTA	CCCATCTGTT	CCACACAAC	TCAGAGTTGC	CACCAGAAAC	2160
ACTACCCTTC	CAAGAGGCGG	TGGTAAAGAC	GGATGCTCGC	CAATTGTTGT	CAAGAAGGGT	2220
CAAGTTGTCA	TGTACACTGT	CATTGGTACC	CACAGAGACC	CAAGTATCTA	CGGTGCCGAC	2280
GCCGACGTCT	TCAGACCAGA	AAGATGTTTC	GAGCCAGAAA	CTAGAAAGTT	GGGCTGGGCA	2340
TATGTTCCAT	TCAATGGTGG	TCCAAGAATC	TGTTTGGGTC	AGCAGTTTGC	CTTGACTGAA	2400
GCTTCATACG	TCAGTGTGAG	ATTGCTCCAA	GAGTTTGGAA	ACTTGTCCTT	GGATCCAAAC	2460
GCTGAGTACC	CACCAAAATT	GCAGAACACC	TTGACCTTGT	CACCTTTTGA	TGGTGCTGAC	2520
GTTAGAATGT	TCTAAGGTTG	CTTATCCTTG	CTAGTGTAT	TTATAGTTTG	TGTATTTAAA	2580
TTGAATCGGC	GATTGATTTT	TCTGGTACTA	ATAACTGTAG	TGGGTTTGA	CCAAAACCGT	2640
TCAAACCTTT	TTTTTTTTTT	TCTTCCCCTT	ACCTTCGTG	CTCGCTCATC	AGCACTGTTT	2700
GAAAACGAAA	AAAGAAAATT	TTTTGTAAAC	AACATTGCCC	AAACTTACCC	AACGTGAACC	2760
ATTATAACCA	AATGAGCGGC	GCTTTCAACT	GGTCACTGGA	GGCATTCCGG	GATATCTACA	2820
ACACCCCTTA	GTTTGAGGAA	GACATTGATT	TAGACACCAT	AGATTTCAGC	GGCATCAAGA	2880
ATGACCTTGT	CCACATTTTG	ACAACCCCAA	CACCACTGGA	AGAATCGCGC	CAGAAACTAG	2940
GCGATGGATC	CAAGCCTGTG	GCCTTGCCCA	ATGGAGACGA	AGTGGAGTTG	AACCAAGCGT	3000
TCCTAGAAGT	TACCACATTA	TTGTGGAACG	AGTTTGACTT	GGACCAATTG	AACGCGGCCG	3060
AGTTGTTATA	CTACGCCGGC	GACATATCCT	ACAAGAAGGG	CACATCAATT	GCCGACAGTG	3120
CCAGATTGTC	TTACTATTTG	AGAGCAAAC	ACATCTTGAA	CATACTGGG	TACTTTATTT	3180
CGAAGCAGCG	ATTGGATGTG	ATAGTCACCG	ACAACAACGC	GTTGTTTGAT	AATATTTTGA	3240
AAAGTTTGA	AAAGATCTAC	AAGTTGATAA	GCGCGTTGAA	CGATATGATT	GACAAGCAAA	3300
AGGTGACAAG	CGACATCAAC	AGTCTAGCAT	TTATCAACTG	CATCAACTAC	TCGAGGGGTC	3360
AACATATTCT	CGCACACGAA	CTTTTGGGAC	TGTTTTGTT	TGGATTGGTT	GACAACTATT	3420
TCAACCAGTT	TGGCTCATT	GACAACTACA	AGAAAGTATT	GGCATTGATA	CTGAAGAACA	3480
TCAGTGATGA	AGATATCTTG	ATCGTACGCT	TCCTCCCATC	GACACTACAA	TTGTTTAAGC	3540
TGGTGTTGGA	TAAGAAAGAC	GACGCCACTG	TTGACCAGTT	CTACAAGTAC	ATCACCTCAA	3600
CAGTGTCGCA	AGACTACAAC	TCCAACATCG	GAGCCACAGC	CAAAGATGAT	ATCGATTTGT	3660
CCAAAGCC						3668

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3826 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

TGGAGTCGCC	AGACTTGCTC	ACTTTTGACT	CCCTTCGAAA	CTCAAAGTAC	GTTCAGGCGG	60
TGCTCAACGA	AACGCTCCGT	ATCTACCCGG	GGGTACCACG	AAACATGAAG	ACAGCTACGT	120
GCAACACGAC	GTTGCCACGC	GGAGGAGGCA	AAGACGGCAA	GGAACCTATC	TTGGTGACAG	180
AGGGACAGTC	CGTTGGGTTG	ATTACTATTG	CCACGCAGAC	GGACCCAGAG	TATTTTGGGG	240
CCGACGCTGG	TGAGTTTAAG	CCGGAGAGAT	GGTTTGATT	AAGCATGAAG	AACTTGGGGT	300
GTAATACTT	GCCGTTCAAT	GCTGGGCCAC	GGACTTGCTT	GGGCGACGAG	TACACTTTGA	360

TTGAAGCGAG	CTACTTGCTA	GTCCGGTTGG	CCCAGACCTA	CCGGGCAATA	GATTTGCGAC	420
CAGGATCGGC	GTACCCACCA	AGAAAGAAGT	CGTTGATCAA	CATGAGTGCT	GCCGACGGGG	480
TGTTTGTAAG	GCTTTATAAG	GATGTAACGG	TAGATGGATA	GTTGTGTAGG	AGGAGCGGAG	540
ATAAATTAGA	TTTGATTTTG	TGTAAGGTTT	TGGATGTCAA	CCTACTCCGC	ACTTTCATGCA	600
GTGTGTGTGA	CACAAGGGTG	TACTACGTGT	GCGTGTGCGC	CAAGAGACAG	CCCAAGGGGG	660
TGGTAGTGTG	TGTTGGCGGA	AGTGCAATGT	ACACAACCGG	TGGGTTCTGG	CCAATGGTGG	720
ACTAAGTGCA	GGTAAGCAGC	GACCTGAAAC	ATTCCTCAAC	GCTTAAGACA	CTGGTGGTAG	780
AGATGCGGAC	CAGGCTATTG	TGTCGTGCT	ACCGGCGGCA	TGGAAAATCA	ACTGCGGGAA	840
GAATAAATTT	ATCCGTAGAA	TCCACAGAGC	GGATAAATTT	GCCACCTCC	ATCATCAACC	900
ACGCCGCCAC	TAACTACATC	ACTCCCTAT	TTTCTCTCTC	TCTCTTTGTC	TTACTCCGCT	960
CCCGTTTCCT	TAGCCACAGA	TACACACCCA	CTGCAAAACG	CAGCAACAAT	TATAAAGATA	1020
CGCCAGGCCC	ACCTTCTTTC	TTTTTCTTCA	CTTTTGTGAC	TGCAACTTTC	TACAATCCAC	1080
CACAGCCACC	ACCACAGCCG	CTATGATTGA	ACAACCTCCTA	GAATATTGGT	ATGTCGTTGT	1140
GCCAGTGTG	TACATCATCA	AACAACCTCT	TGCATACACA	AAGACTCGCG	TCTTGATGAA	1200
AAAGTTGGGT	GCTGCTCCAG	TCACAAACAA	ATTGTACGAC	AACGCTTTTCG	GTATCGTCAA	1260
TGGATGGAAG	GCTCTCCAGT	TCAAGAAAGA	GGGCAGGGCT	CAAGAGTACA	ACGATTACAA	1320
GTTTGACCAC	TCCAAGAACC	CAAGCGTGGG	CACCTACGTC	AGTATTCTTT	TCGGCACCAG	1380
GATCGTCGTG	ACCAAAGATC	CAGAGAATAT	CAAAGCTATT	TTGGCAACCC	AGTTTGGTGA	1440
TTTTTCTTTG	GGCAAGAGGC	ACACTCTTTT	TAAGCCTTTG	TTAGGTGATG	GGATCTTCAC	1500
ATTGACCGGC	GAAGGCTGGA	AGCACAGCAG	AGCCATGTTG	AGACCACAGT	TTGCCAGAGA	1560
ACAAGTTGCT	CATGTGACGT	CGTTGGAACC	ACACTTCCAG	TTGTTGAAGA	AGCATATTCT	1620
TAAGCACAAG	GGTGAATACT	TTGATATCCA	GGAAATGTTT	TTAGATTTA	CCGTTGATTC	1680
GGCCACGGAG	TTCTTATTTG	GTGAGTCCGT	GCACTCCTTA	AAGGACGAAT	CTATTGGTAT	1740
CAACCAAGAC	GATATAGATT	TTGCTGGTAG	AAAGGACTTT	GCTGAGTCGT	TCAACAAAGC	1800
CCAGGAATAC	TTGGCTATTA	GAACCTTGGT	GCAGACGTTT	TACTGGTTGG	TCAACAACAA	1860
GGAGTTTAGA	GACTGTACCA	AGCTGGTGCA	CAAGTTCACC	AACACTATG	TTCAGAAAGC	1920
TTTGGATGCT	AGCCCAGAAG	AGCTTGAAAA	GCAAAGTGGG	TATGTGTTCT	TGTACGAGCT	1980
TGTCAGACAG	ACAAGAGACC	CCAATGTGTT	GCGTGACCAG	TCTTTGAACA	TCTTGTGGC	2040
CGGAAGAGAC	ACCACTGCTG	GGTTGTGTC	GTTTGCTGTC	TTGAGTTGG	CCAGACACCC	2100
AGAGATCTGG	GCCAAAGTTGA	GAGAGGAAAT	TGAACAACAG	TTGGTCTTG	GAGAAGACTC	2160
TCGTGTTGAA	GAGATTACCT	TTGAGAGCTT	GAAGAGATGT	GAGTACTTGA	AAGCGTTCCT	2220
TAATGAAACC	TTGCGTATTT	ACCCAAGTGT	CCCAAGAAAC	TTCAGAATCG	CCACCAAGAA	2280
CACGACATTG	CCAAGGGGCG	GTGGTTCAGA	CGGTACCTCG	CCAATCTTGA	TCCAAAGGGG	2340
AGAACTGTG	TCGTATGGTA	TCAACTCTAC	TCATTTGGAC	CCTGTCTATT	ACGGCCCTGA	2400
TGCTGCTGAG	TTCAGACCAG	AGAGATGGTT	TGAGCCATCA	ACCAAAAAGC	TCGGCTGGGC	2460
TTACTTGCCA	TTCAACGGTG	GTCCAAGAAT	CTGTTTGGGT	CAGCAGTTTG	CCTTGACGGA	2520
AGCTGGCTAT	GTGTTGGTTA	GATTGGTGCA	AGAGTTCCTC	CACGTTAGGC	TGGACCCAGA	2580
CGAGGTGTAC	CCGCCAAAGA	GGTTGACCAA	CTTGACCATG	TGTTTGACAG	ATGGTGCTAT	2640
TGTCAGATTT	GACTAGCGGC	GTGGTGAATG	CGTTTGATTT	TGTAGTTTCT	GTTTGACGTA	2700
ATGAGATAAC	TATTAGATA	AGGCGAGTGG	ATGTACGTTT	TGTAAGAGTT	TCCTTACAAC	2760
CTTGTGGGG	TGTGTGAGGT	TGAGGTTGCA	TCTTGGGGAG	ATTACACCTT	TTGCAGCTCT	2820
CCGTATACAC	TTGTACTCTT	TGTAACCTCT	ATCAATCATG	TGGGGGGGGG	GGTTCATTGT	2880
TTGGCCATGG	TGGTGCATGT	TAAATCCGCC	AACCTACCAA	TCTCACATGA	AACCTAAGCA	2940
CACTAAAAAA	AAAAAAGATG	TTGGGGGAAA	ACTTTGGTTT	CCCTTCTTAG	TAATTAAACA	3000
CTCTCACTCT	CACTCTCACT	CTCTCCACTC	AGACAAACCA	ACCACCTGGG	CTGCAGACAA	3060
CCAGAAAAAA	AAAGAACAAA	ATCCAGATAG	AAAAACAAAG	GGCTGGACAA	CCATAAATAA	3120
ACAACTTAGG	GTCTACTCCA	TCTTCCACTG	TTTCTTCTTC	TTAGACTTGA	GCTAACAAAC	3180
AACTCACTTC	ACCATGGATT	ACGCAGGCAT	CACGCGTGGC	TCCATCAGAG	GCGAGGCCTT	3240
GAAGAACTC	GCAGAAATGA	CCATCCAGAA	CCAGCCATCC	AGCTTGAAAG	AAATCAACAC	3300
CGGCATCCAG	AAGGACGACT	TTGCCAAGTT	GTTGCTCGCC	ACCCCGAAAA	TCCCAACCAA	3360
GCACAAAGTT	AACGGCAACC	ACGAATTGTC	TGAGGTCGCC	ATTGCCAAAA	AGGAGTACGA	3420
GGTGTGATT	GCCTTGAGCG	ACGCCACAAA	AGACCAATC	AAAGTGACCT	CCCAGATCAA	3480
GATCTTGATT	GACAAGTTCA	AGGTGTACTT	GTTTGAGTTG	CCTGACCAGA	AGTTCCTCTA	3540
CTCCATCGTG	TCCAACCTCCG	TCAACATCGC	CCCCTGGACC	TTGCTCGGGG	AGAAGTTGAC	3600
CACGGGCTTG	ATCAACTTGG	CCTTCCAGAA	CAACAAGCAG	CACTTGGACG	AGGTCAATTGA	3660
CATCTTCAAC	GAGTTCATCG	ACAAGTTCTT	TGGCAACACG	GAGCCGCAAT	TGACCAACTT	3720

CTTGACCTTG	TGCGGTGTGT	TGGACGGGTT	GATTGACCAT	GCCAACTTCT	TGAGCGTGTC	3780
CTCGCGGACC	TTCAAGATCT	TCTTGAACCT	GGACTCGTAT	GTGGAC		3826

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3910 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

TTACAATCAT	GGAGCTCGCT	AGGAACCCAG	ATGCTGCGGA	GAAGCTCCGC	GAAGAGGTCA	60
ACACGAACTT	TGGCATGGAG	TCGCCAGACT	TGCTCACTTT	TGACTCTCTT	AGAAGCTCAA	120
AGTACGTTCA	GGCGGTGCTC	AACGAAACGC	TTCGTATCTA	CCCGGGGGTG	CCACGAAACA	180
TGAAGACAGC	TACGTGCAAC	ACGACGTTGC	CGCGTGGAGG	AGGCAAAGAC	GGTAAGGAAC	240
CTATTTTGGT	GCAGAAGGGC	CAGTCCGTTG	GGTTGATTAC	TATTGCCACG	CAGACGGACC	300
CAGAGTATTT	TGGGGCAGAT	GCTGGTGAGT	TCAAACCGGA	GAGATGGTTT	GATTCAAGCA	360
TGAAGAACTT	GGGGTGTAAG	TACTTGCCGT	TCAATGCTGG	GCCCCGGACT	TGTTTGGGGC	420
AGCAGTACAC	TTTGATTGAA	GCGAGCTATT	TGCTAGTCAG	GTTGGCGCAG	ACCTACCGGG	480
TAATCGATTT	GCTGCCAGGG	TCGGCGTACC	CACCAAGAAA	GAAGTCGTTG	ATCAATATGA	540
GTGCTGCCGA	TGGGGTGTTT	GTAAAGTTTC	ACAAGGATCT	AGATGGATAT	GTAAGGTGTG	600
TAGGAGGAGC	GGAGATAAAT	TAGATTTGAT	TTTGTGTAAG	GTTTAGCACG	TCAAGCTACT	660
CCGCACTTTG	TGTGTAGGGA	GCACATACTC	CGTCTGCGCC	TGTGCCAAGA	GACGGCCAG	720
GGGTAGTGTG	TGGTGGTGGA	AGTGCATGTG	ACACAATACC	CTGGTTCTGG	CCAATTGGGG	780
ATTTAGTGTA	GGTAAGCTGC	GACCTGAAAC	ACTCCTCAAC	GCTTGAGACA	CTGGTGGGTA	840
GAGATGCGGG	CCAGGAGGCT	ATTCTTGTCG	TGCTACCCGT	GCACGGAAAA	TCGATTGAGG	900
GAAGAACAAA	TTTATCCGTG	AAATCCACAG	AGCGGATAAA	TTTGTACAT	TGCTGCGTTG	960
CCCACCCACA	GCATTCTCTT	TTCTCTCTCT	TTGTCTTACT	CCGCTCCTGT	TTCTTTATCC	1020
AGAAATACAC	ACCAACTCAT	ATAAAGATAC	GCTAGCCCAG	CTGTCTTTCT	TTTTCTTCAC	1080
TTTTTTTGGT	GTGTTGCTTT	TTTGGCTGCT	ACTTTCTACA	ACCACCACCA	CCACCACCAC	1140
CATGATTGAA	CAAATCCTAG	AATATTGGTA	TATTGTTGTG	CCTGTGTTGT	ACATCATCAA	1200
ACAACCTCATT	GCCTACAGCA	AGACTCGCGT	CTTGATGAAA	CAGTTGGGTG	CTGCTCCAAT	1260
CACAAACCAG	TTGTACGACA	ACGTTTTTCG	TATCGTCAAC	GGATGGAAGG	CTCTCCAGTT	1320
CAAGAAAGAG	GGCAGAGCTC	AAGAGTACAA	CGATCACAA	TTTGACAGCT	CCAAGAACCC	1380
AAGCGTCGGC	ACCTATGTCA	GTATTCTTTT	TGGCACCAAG	ATTGTCGTGA	CCAAGGATCC	1440
AGAGAAATATC	AAAGCTATTT	TGGCAACCCA	GTTTGGCGAT	TTTTCTTTGG	GCAAGAGACA	1500
CGCTCTTTT	AAACCTTTGT	TAGGTGATGG	GATCTTCACC	TTGGACGGCG	AAGGCTGGAA	1560
GCATAGCAGA	TCCATGTTAA	GACCACAGTT	TGCCAGAGAA	CAAGTTGCTC	ATGTGACGTC	1620
GTTGGAACCA	CACTTCCAGT	TGTTGAAGAA	GCATATCCTT	AAACACAAGG	GTGAGTACTT	1680
TGATATCCAG	GAATTGTTCT	TTAGATTTAC	TGTCGACTCG	GCCACGGAGT	TCTTATTGG	1740
TGAGTCCGTG	CACTCCTTAA	AGGACGAAAC	TATCGGTATC	AACCAAGACG	ATATAGATTT	1800
TGCTGGTAGA	AAGGACTTTG	CTGAGTCGTT	CAACAAAGCC	CAGGAGTATT	TGTCTATTAG	1860
AATTTTGGTG	CAGACCTTCT	ACTGGTTGAT	CAACAACAAG	GAGTTTAGAG	ACTGTACCAA	1920
GCTGGTGAC	AAGTTTACCA	ACTACTATGT	TCAGAAAGCT	TTGGATGCTA	CCCCAGAGGA	1980
ACTTGAAAAG	CAAGGCGGGT	ATGTGTTCTT	GTATGAGCTT	GTCAAGCAGA	CGAGAGACCC	2040
CAAGGTGTTG	CGTGACCAGT	CTTTGAACAT	CTTGTTGGCA	GGAAGAGACA	CCACTGCTGG	2100
GTTGTTGTCC	TTTGTGTGT	TTGAGTTGGC	CAGAAACCCA	CACATCTGGG	CCAAGTTGAG	2160
AGAGGAAATT	GAACAGCAGT	TTGGTCTTGG	AGAAGACTCT	CGTGTGGAAG	AGATTACCTT	2220
TGAGAGCTTG	AAGAGATGTG	AGTACTTGAA	AGCGTTCCTT	AACGAAACCT	TGCGTGTTTA	2280
CCCAAGTGTC	CCAAGAAACT	TCAGAAATCG	CACCAAGAAT	ACAACATTGC	CAAGGGGTGG	2340
TGGTCCAGAC	GGTACCCAGC	CAATCTTGAT	CCAAAGGGGA	GAAGGTGTGT	CGTATGTGAT	2400
CAACTCTACC	CACTTAGATC	CTGTCTATTA	TGGCCCTGAT	GCTGCTGAGT	TCAGACCAGA	2460
GAGATGGTTT	GAGCCATCAA	CCAGAAAGCT	CGGCTGGGCT	TACTTGCCAT	TCAACGGTGG	2520
GCCACGAATC	TGTTTGGGTC	AGCAGTTTGC	CTTGACCGAA	GCTGGTTACG	TTTGGGTGAG	2580
ATTGGTGCAA	GAGTTCTCCC	ACATTAGGCT	GGACCCAGAT	GAAGTGTATC	CACCAAGAG	2640
GTTGACCAAC	TTGACCATGT	GTTTGCAGGA	TGGTGTCTATT	GTCAAGTTTG	ACTAGTACGT	2700

ATGAGTGCCT	TTGATTTTGT	AGTTTCTGTT	TGCAGTAATG	AGATAACTAT	TCAGATAAGG	2760
CGGGTGGATG	TACGTTTTGT	AAGAGTTTCC	TTACAACCCT	GGTGGGTGTG	TGAGGTTGCA	2820
TCTTAGGGAG	AGATAGCACC	TTTTCAGCT	CTCCGTATAC	AGTTTACTC	TTTGTAACCT	2880
ATGCCAATCA	TGTGGGGATT	CATTGTTTGC	CCATGGTGGT	GCATGCAAAA	TCCCCCAAC	2940
TACCCAATCT	CACATGAAAC	TCAAGCACAC	TAGAAAAAAA	AGATGTTGCG	TGGGTTCTTT	3000
TGATGTTGGG	GAAAACTTTC	GTTTCCTTTC	TCAGTAATTA	AACGTTCTCA	CTCAGACAAA	3060
CCACCTGGGC	TGCAGACAAC	CAGAAAAAAC	AAAATCCAGA	TAGAAGAAGA	AAGGGCTGGA	3120
CAACCATAAA	TAAACAACCT	AGGGTCCACT	CCATCTTTCA	CTTCTTCTTC	TTCAGACTTA	3180
TCTAACAAAC	GACTCACTTC	ACCATGGATT	ACGCAGGTAT	CACGGGTGGG	TCCATCAGAG	3240
GCGAAGCCTT	GAAGAACTC	GCCGAGTTGA	CCATCCAGAA	CCAGCCATCC	AGCTTGAAAG	3300
AAATCAACAC	CGGCATCCAG	AAGGACGACT	TTGCCAAGTT	GTTGTCTTCC	ACCCCGAAAA	3360
TCCACACCAA	GCACAAGTTG	AATGGCAACC	ACGAATTGTC	CGAAGTCGCC	ATTGCCAAAA	3420
AGGAGTACGA	GGTGTGATT	GCCTTGAGCG	ACGCCACGAA	AGAACCACATC	AAAGTCACCT	3480
CCCAGATCAA	GATCTTGATT	GACAAGTTCA	AGGTGTACTT	GTTTGAGTTG	CCCGACCAGA	3540
AGTTCTCCTA	CTCCATCGTG	TCCAACCTCCG	TTAACATTGC	CCCCTGGACC	TTGCTCGGTG	3600
AGAAGTTGAC	CACGGGCTTG	ATCAACTTGG	CGTTCCAGAA	CAACAAGCAG	CACTTGGACG	3660
AAGTCATCGA	CATCTTCAAC	GAGTTCATCG	ACAAGTTCTT	TGGCAACACA	GAGCCGCAAT	3720
TGACCAACTT	CTTGACCTTG	TCCGGTGTGT	TGGACGGGTT	GATTGACCAT	GCCAACCTCT	3780
TGAGCGTGTC	CTCCAGGACC	TTCAAGATCT	TCTTGAACCT	GGACTCGTTT	GTGGACAACCT	3840
CGGACTTCTT	GAACGACGTG	GAGAACTACT	CCGACTTTTT	GTACGACGAG	CCGAACGAGT	3900
ACCAGAACTT						3910

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3150 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

GAATTCCTTG	GATCTAATTC	CAGCTGATCT	TGCTAATCCT	TATCAACGTA	GTTGTGATCA	60
TTGTTTGTCT	GAATTATACA	CACCAGTGGA	AGAATATGGT	CTAATTTGCA	CGTCCCCTG	120
GCATTGTGTG	TTTGTGGGGG	GGGGGGGGTG	CACACATTTT	TAGTGCCATT	CTTTGTTGAT	180
TACCCCTCCC	CCCTATCATT	CATTCCCA	GGATTAGTTT	TTTCTCTACT	GGAATTCGCT	240
GTCCACCTGT	CAACCCCCC	CCCCCCCCC	CCCACTGCCC	TACCCTGCCC	TGCCCTGCAC	300
GTCCTGTGTT	TTGTGCTGTG	TCTTTCCAC	GCTATAAAAG	CCCTGGCGTC	CGGCCAAGGT	360
TTTTCCACCC	AGCCAAAAAA	ACAGTCTAAA	AAATTTGGTT	GATCCTTTTT	GGTTGCAAGG	420
TTTTCCACCA	CCACTTCCAC	CACCTCAACT	ATTGGAACAA	AAGATGCTCG	ATCAGATCTT	480
ACATTACTGG	TACATTGTCT	TGCCATTGTT	GGCCATTATC	AACCAGATCG	TGGCTCATGT	540
CAGGACCAAT	TATTTGATGA	AGAAATTGGG	TGCTAAGCCA	TTCAACACG	TCCAACGTGA	600
CGGGTGGTTG	GGCTTCAAAT	TCGGCCGTGA	ATTCTCCTAA	GCAAAAAGTG	CTGGGAGACT	660
GGTTGATTTA	ATCATCTCCC	GTTTCCACGA	TAATGAGGAC	ACTTTCTCCA	GCTATGCTTT	720
TGGCAACCAT	GTGGTGTTC	CCAGGGACCC	CGAGAATATC	AAGGCGCTTT	TGGCAACCCA	780
GTTTGGTGAT	TTTTCAATTG	GCAGCAGGGT	CAAGTTCTTC	AAACCATTAT	TGGGGTACGG	840
TATCTTCACA	TTGGACGCCG	AAGGCTGGAA	GCACAGCAGA	GCCATGTTGA	GACCACAGTT	900
TGCCAGAGAA	CAAGTTGCTC	ATGTGACGTC	GTTGGAACCA	CACCTCCAGT	TGTTGAAGAA	960
GCATATCCTT	AAACACAAGG	GTGAGTACTT	TGATATCCAG	GAATTGTTCT	TTAGATTTC	1020
TGTCGACTCG	GCCACGAGT	TCTTATTTGG	TGAGTCCGTG	CACCTCTTAA	AGGACGAGGA	1080
AATTGGCTAC	GACACGAAAG	ACATGTCTGA	AGAAAGACGC	AGATTGCGG	ACGCGTTCAA	1140
CAAGTCGCAA	GTCTACGTGG	CCACCAGAGT	TGCTTTACAG	AACTTGTA	GGTTGGTCAA	1200
CAACAAGAG	TTCAAGGAGT	GCAATGACAT	TGTCACAAAG	TTTACCAACT	ACTATGTTCA	1260
GAAAGCCTTG	GATGCTACCC	CAGAGGAACT	TGAAAAGCAA	GGCGGGTATG	TGTTCTTGTA	1320
TGAGCTTGTC	AAGCAGACGA	GAGACCCCAA	GGTGTGCGT	GACCAGTCTT	TGAACATCTT	1380
GTTGGCAGGA	AGAGACACCA	CTGCTGGGTT	GTTGTCCTTT	GCTGTGTTTG	AGTTGGCCAG	1440
AAACCCACAC	ATCTGGGCCA	AGTTGAGAGA	GGAAATTGAA	CAGCAGTTTG	GTCTTGAGGA	1500
AGACTCTCGT	GTTGAAGAGA	TTACCTTTGA	GAGCTTGAAG	AGATGTGAGT	ACTTGAAGGC	1560

CGTGTGTAAC	GAAACTTTGA	GATTACACCC	AAGTGTCCCA	AGAAACGCAA	GATTTGCGAT	1620
TAAAGACACG	ACTTTACCAA	GAGGCGGTGG	CCCCAACGGC	AAGGATCCTA	TCTTGATCAG	1680
GAAGGATGAG	GTGGTGCAGT	ACTCCATCTC	GGCAACTCAG	ACAAATCCTG	CTTATTATGG	1740
CGCCGATGCT	GCTGATTTTA	GACCGGAAAG	ATGGTTTGAA	CCATCAACTA	GAAACTTGGG	1800
ATGGGCTTTC	TTGCCATTCA	ACGGTGGTCC	AAGAATCTGT	TTGGGACAAC	AGTTTGCTTT	1860
GACTGAAGCC	GGTTACGTTT	TGGTTAGACT	TGTTACAGGAG	TTTCCAAACT	TGTCACAAGA	1920
CCCCGAAACC	AAGTACCCAC	CACCTAGATT	GGCACACTTG	ACGATGTGCT	TGTTTGACGG	1980
TGCACACGTC	AAGATGTCAT	AGGTTTCCCC	ATACAAGTAG	TTCAGTAATT	ATACACTGTT	2040
TTTACTTTCT	CTTCATACCA	AATGGACAAA	AGTTTTAAGC	ATGCCTAACA	ACGTGACCGG	2100
ACAATTGTGT	CGCACTAGTA	TGTAACAATT	GTAAAAATAG	TGTACACTAA	TTTGTGGTGG	2160
CCGGAGATAA	ATTACAGTTT	GGTTTTGTGT	AAACTCGCGG	ATATCTCTGG	CAGTTTCTCT	2220
TCTCCGCAGC	AGCTTTGCCA	CGGGTTTGCT	CTGGGGCCAA	CAAAATCAAA	AGGGGGAGAA	2280
ACTTAACACC	CCTTATCTCT	CCACTCTAGG	TTGTAGCTCT	TGTGGGGATG	CAATTGTCTG	2340
ACGTTTTTTA	TGTTTTGTCT	AGACTTTGAT	GATTACGTTG	GATTTCTTAT	GTCTGAGCGG	2400
TGCTTGAAAG	AAGTGTCAAA	ATGTGACAGG	CGACGCTATT	CGACATGAAC	GCGAAAGGGT	2460
TATTTGCATC	AATACGAGGG	GCTGACTCTA	GTCTAGGATG	GCAGTCCTAG	GTTGCAAACA	2520
TGTTGCACCA	TATCCCTCCT	GGAGTTGGTC	GACCTCGCCT	ACGCCACCCT	CAGCGATCGG	2580
CACTTTCCGT	TGTTCAATAT	TTCTCCTTCC	CATTGTTCCA	GGGTTATCA	ACAACGTTGC	2640
CGGCCTCCTC	CCCAAATTAC	AAGAAAATA	AATTGTGCGA	CGGCACCGAT	CTGTCAAAGA	2700
TACAGATAAA	CCTTAAATCT	GCAAAAACAA	GACCCCTCCC	CATAGCCTAG	AAGCACCAGC	2760
AAGATGATGG	AGCAACTCCT	CCAGTACTGG	TACATCGCAC	TCTCTGTATG	GTTTCATCCT	2820
CGCTACTTGG	CTTCCACCGC	ACGAGCCGTC	TACTTGCGCC	ACAAGCTCGG	CGCGCGCCCA	2880
TTACAGCACA	CCAGTACGA	CGGCTGGTAT	GGGTTCAAGT	TTGGGCGGGA	GTTTCTCAAG	2940
GCGAAGAAGA	TCGGGCGGCA	GACGGACTTG	GTGCATGCGC	GGTTCGTTGG	CGGCATGGAC	3000
ACCTTCTCGA	GCTACACTTT	CGGCATCCAT	ATCATCCTTA	CCGCGGACCC	GGAGAACATC	3060
AAGGCGGTCT	TGGCGACGCA	GTTTCGATGAC	TTCTCGCTCG	GTGCAGGAT	CAGGTTCTTG	3120
AAGCCGTTGT	TGGGGTATGG	GATATTCACG				3150

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3579 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

AAAACCGATA	CAAGAAGAAG	ACAGTCAACA	AGAACGTTAA	TGTCAACCAG	GCGCCAAGAA	60
GACGGTTTGG	CGGACTTGGA	AGAATGTGGC	ATTTGCCCAT	GATGTTTATG	TTCTGGAGAG	120
GTTTTTCAAG	GAATCGTCAT	CCTCCGCCAC	CACAAGAACC	ACCAGTTAAC	GAGATCCATA	180
TTCAACAACC	ACCGCAAGGT	GACATGCTC	AACAACAACA	GCAACAACAA	CAACCCCCAC	240
AGAACAGTG	GAATAATGCC	AGTCAACAAA	GAGTGGTGAC	AGACGAGGGA	GAAAACGCAA	300
GCAACAGTGG	TTCTGATGCA	AGATCAGCTA	CACCGTTCA	TCAGGAAAAG	CAGGAGCTCC	360
CACCAACATA	TGCCCATCAC	GAGCAACACC	AGCAGGTTAG	TGTATAGTAG	TCTGTAGTTA	420
AGTCAATGCA	ATGTACCAAT	AAGACTATCC	CTTCTTACAA	CCAAGTTTTC	TGCCGCGCCT	480
GTCTGGCAAC	AGATGCTGGC	CGACACACTT	TCAACTGAGT	TTGGTCTAGA	ATTCTTGCAC	540
ATGCACGACA	AGGAAACTCT	TACAAAGACA	ACACTTGTGC	TCTGATGCCA	CTTGATCTTG	600
CTAAGCCTTA	TCAACGTAAT	TGAGATCATT	GTTTGTCTGA	ATTATACACA	CCAGTGGAAG	660
AATCTGGTCT	AATCTGCACG	CCTCATGGGC	ATTGTGTGTT	TTGGGGGGGG	GGGGGGGGGT	720
GCACACATTT	TTAGTGCAGAA	TGTTTGTGTT	CTGGTTCCCC	CTCCCCCTC	CCCCCTATCA	780
TGCCCCACAG	ATTAGTTTTT	TCCTCACTGG	AATTGCTGTG	CCACCTGTCA	ACCCCTCAC	840
TGCCCTGCCC	TGCCCTGCAC	GCCCTGTGTT	TTGTGCTGTG	GCACTCCAC	GCTATAAAAG	900
CCCTGGCGTA	CGGCCAAGGT	TTTTCTCAC	AGCCAAAAAA	AAATTTGGCT	GATCCTTTTG	960
GGCTGCAAGG	TTTTTCACCA	CCACCACCAC	CACCACCTCA	ACTATTCAAA	CAAAGGATGC	1020
TCGACCAGAT	CTTCCATTAC	TGGTACATTG	TCTTGCCATT	GTTGGTCATT	ATCAAGCAGA	1080
TCGTGGCTCA	TGCCAGGACC	AATTATTTGA	TGAAGAAGTT	GGGCGCTAAG	CCATTACAC	1140
ATGTCCAAC	AGACGGGTGG	TTTGGCTTCA	AATTTGGCCG	TGAATTCCTC	AAAGCTAAAA	1200

GTGCTGGGAG	GCAGGTTGAT	TTAATCATCT	CCCGTTTCCA	CGATAATGAG	GACACTTTCT	1260
CCAGCTATGC	TTTGGCAAC	CATGTGGTGT	TCACCAGGGA	CCCCGAGAAT	ATCAAGGCGC	1320
TTTTGGCAAC	CCAGTTTGGT	GATTTTTTCAT	TGGAAGCAG	GGTCAAATTC	TTCAAACCAT	1380
TGTTGGGGTA	CGGTATCTTC	ACCTTGGACG	GCGAAGGCTG	GAAGCACAGC	AGAGCCATGT	1440
TGAGACCACA	GTTTGCCAGA	GAGCAAGTGT	CTCATGTGAC	GTCGTTGGAA	CCACATTTCC	1500
AGTTGTTGAA	GAAGCATATT	CTTAAGCACA	AGGGTGAATA	CTTTGATATC	CAGGAATTGT	1560
TCTTTAGATT	TACCGTTGAT	TCAGCGACGG	AGTTCTTATT	TGGTGAGTCC	GTGCACTCCT	1620
TAAGGGACGA	GGAAATTGGC	TACGATACGA	AGGACATGGC	TGAAGAAAGA	CGCAAATTGT	1680
CCGACGCGTT	CAACAAGTCG	CAAGTCTATT	TGTCCACCAG	AGTTGCTTTA	CAGACATTGT	1740
ACTGTTGGT	CAACAACAAA	GAGTTCAAGG	AGTGCAACGA	CATTGTCCAC	AAGTTCCACCA	1800
ACTACTATGT	TCAGAAAGCC	TTGGATGCTA	CCCCAGAGGA	ACTTGAAAAA	CAAGGCGGGT	1860
ATGTGTTCTT	GTACGAGCTT	GCCAAGCAGA	CGAAAGACCC	CAATGTGTTG	CGTGACCAGT	1920
CTTTGAACAT	CTTGTTGGCT	GGAAGGGACA	CCACTGCTGG	GTTGTTGTCC	TTTGCTGTGT	1980
TTGCTTGGC	CAGGAACCCA	CACATCTGGG	CCAAGTTGAG	AGAGGAAATT	GAATCACT	2040
TTGGGCTGGG	TGAGGACTCT	CGTGTGAAG	AGATTACCTT	TGAGAGCTTG	AAGAGATGTG	2100
AGTACTTGAA	AGCCGTGTTG	AACGAAACGT	TGAGATTACA	CCCAAGTGTC	CCAAGAAACG	2160
CAAGATTGTC	GATTAAAGAC	ACGACTTTAC	CAAGAGGCGG	TGGCCCCAAC	GGCAAGGATC	2220
CTATCTTGAT	CAGAAAGAAT	GAGGTGGTGC	AATACTCCAT	CTCGGCAACT	CAGACAAATC	2280
CTGCTTATTA	TGGCGCCGAT	GCTGCTGATT	TTAGACCGGA	AAGATGGTTT	GAGCCATCAA	2340
CTAGAAACTT	GGGATGGGCT	TACTTGCCAT	TCAACGGTGG	TCCAAGAATC	TGCTTGGGAC	2400
AACAGTTTGC	TTTGACCGAA	GCCGGTTACG	TTTTGGTTAG	ACTTGTTTCA	GAATTCCTTA	2460
GCTTGTCACA	GGACCCCGAA	ACTGAGTACC	CACCACCTAG	ATTGGCACAC	TTGACGATGT	2520
GCTTGTGTTA	CGGGGCATAC	GTCAAGATGC	AATAGGTTTT	GGTTTGACTT	TGTTTCCATA	2580
TGCAAGTAGT	TCAGTAATTA	CACACTAATT	TGTGGTGGCC	GGCGATAAAT	TACCGTTTGG	2640
TTTTGTGTAA	AAATTCGGAC	ATCTCTGGTG	GTTTCCCTTC	TCCGCAGCAG	CTTTGCCACG	2700
GGTTTGCTCT	CGGGCCAAAC	AATTCGAAAG	GGGGGGGGGG	GGGGGAGAAA	GTTAACACCC	2760
CCTGTTCCCA	CCGTAGGCTG	TAGCTCTTGT	GGGGGGATGT	AATTGTCGTA	CGTTTTTCATG	2820
TTTGGCCCAG	ACTTTGATGA	TTACGTAGGC	TTTCTTATGT	CTAAGGCGTG	CTTGACACAA	2880
GTGTCAAAAG	GTGACAGGCG	ACGTTATTCG	ACATGAACGC	AAAAGGGTAA	TTTGATCGCA	2940
TACGAGGGGT	TGCCTCTGGT	CTAAGAAGGA	CCCCCAGGT	TGCAAAACATG	TTGCACTGCA	3000
TCCCCTCAG	AGTTGGTCGA	CCACGCCTAC	GCTTACCCTC	AGCGATCGGC	ACTTTCCGTT	3060
GCTCAATATT	TCTCTCCCCC	CTGCTTCCCC	CCATTGTTCC	AGGGATTATC	AACAACGTTG	3120
CCGTCTCCT	CTCCCCCCCC	TCCCCCAGT	TATGTACAAG	AAAATTAAAT	TGTCGCACGG	3180
CACCGATACG	TCAAAGATAC	AGAGAAACCT	TAATCCCTCC	CATAGCCTAG	AAGCATCAAA	3240
AAGATGATTG	AGCAACTCCT	CCAGTACTGG	TACATTGCAC	TCCCTGTATG	GTTTATTCTC	3300
CGCTACGTGG	CTTCCCACGC	ACGAACCATC	TACTTGCGCC	ACAAGCTCGG	CGCGGCGCGG	3360
TTCAAGCACA	CCCAGTACGA	CGGATGGTAT	GGGTTCAAGT	TTGGGCGGGA	GTTTCTCAAG	3420
GCGAAGAAGA	TTGGAAGGCA	GACGGAATTG	GTGCATGCGC	GGTTCCGTGG	AGGGGGCATG	3480
GATACTTTCT	CGAGCTATAC	TTTCGGCATC	CATATCATTC	TTACTCGGGA	CCCGGAGAAC	3540
ATCAAGGCGG	TCTTGCGGAC	GCAGTTGAT	GACTTTTCTG			3579

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3348 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GATGTGGTGC	TTGATTCTC	GAGACATC	CTGTGAGGT	GCCATGAATC	TGTACCTGTC	60
TGTAAGCACA	GGGAAGTGT	TCAACACCTT	ATTGCATATT	CTGTCTATTG	CAAGCGTGTG	120
CTGCAACGAT	ATCTGCCAAG	GTATATAGCA	GAACGTGCTG	ATGGTTCTCT	CGGTCATATT	180
CTGTTGGTAG	TTCTGCAGGT	AAATTTGGAT	GTCAGGTAGT	GGAGGGAGGT	TTGTATCGGT	240
TGTGTTTTCT	TCTTCTCTC	TCTCTGATC	AACCTCCACG	TCTCCTTCGG	GTTCTGTGTC	300
TGTGCTCTGAG	TCGTACTGTT	GGATTAAAGT	CATCGCATGT	GTGAAAAAAA	GTAGCGCTTA	360
TTTAGACAAC	CAGTTCGTTG	GCGGGTATC	AGAAATAGTC	TGTTGTGCAC	GACCATGAGT	420

ATGCAACTTG	ACGAGACGTC	GTTAGGAATC	CACAGAATGA	TAGCAGGAAG	CTTACTACGT	480
GAGAGATTCT	GCTTAGAGGA	TGTTCTCTTC	TTGTTGATTG	CATTAGGTGG	GTATCATCTC	540
CGGTGGTGAC	AACTTGACAC	AAGCAGTTCC	GAGAACCACC	CACAACAATC	ACCATTCCAG	600
CTATCACTTC	TACATGTCAA	CCTACGATGT	ATCTCATCAC	CATCTAGTTT	CTTGGCAATC	660
GTTTATTTGT	TATGGGTCAA	CATCCAATAC	AACTCCACCA	ATGAAGAAGA	AAAACGGAAA	720
GCAGAATACC	AGAAATGACAG	TGTGAGTTCC	TGACCATTGC	TAATCTATGG	CTATATCTAG	780
TTTGCTATCG	TGGGATGTGA	TCTGTGTCGT	CTTCATTTGC	GTTTGTGTTT	ATTTCCGGTA	840
TGAATATTGT	TATACTAAAT	ACTTGATGCA	CAAACATGGC	GCTCGAGAAA	TCGAGAAATG	900
GATCAACGAT	GGGTTCTTTG	GGTTCGCTT	ACCTTTGCTA	CTCATGCGAG	CCAGCAATGA	960
GGGCCGACTT	ATCGAGTTCA	GTGTCAAGAG	ATTCGAGTCG	GCGCCACATC	CACAGAACAA	1020
GACATTGGTG	AACCGGGCAT	TGACGCTTCC	TGTGATACTC	ACCAAGGACC	CAGTGAATAT	1080
CAAAGCGATG	CTATCGACCC	AGTTTGATGA	CTTTCCCTT	GGGTGAGAC	TACACCAGTT	1140
TGCGCCGTTG	TTGGGGAAG	GCATCTTTAC	TTTGGACGGC	CCAGAGTGGG	AGCAGAGCCG	1200
ATCTATGTTG	CGTCCGCAAT	TTGCCAAGA	TCGGGTTTCT	CATATCCTGG	ATCTAGAACC	1260
GCATTTTGTG	TTGCTTCGGA	AGCACATTGA	TGGCCACAAT	GGAGACTACT	TCGACATCCA	1320
GGAGCTCTAC	TTCCGGTTCT	CGATGGATGT	GGCGACGGGG	TTTTTGTGTT	GCGAGTCTGT	1380
GGGGTCGTTG	AAAGACGAAG	ATGCGAGGTT	CCTGGAAGCA	TTCAATGAGT	CGCAGAAGTA	1440
TTTGCCAAT	AGGGCAACGT	TGCACGAGTT	GTACTTTCTT	TGTGACGGGT	TTAGGTTTCG	1500
CCAGTACAAC	AAGGTTGTGC	GAAAGTTCTG	CAGCCAGTGT	GTCCACAAGG	CGTTAGATGT	1560
TGCACCGGAA	GACACCAGCG	AGTACGTGTT	TCTCCGCGAG	TTGGTCAAAC	ACACTCGAGA	1620
TCCCGTTGTT	TTACAAGACC	AAGCGTTGAA	CGTCTTGCTT	GCTGGACGCG	ACACCACCGC	1680
GTCGTTATTA	TCGTTTGCAA	CATTTGAGCT	AGCCCGGAAT	GACCACATGT	GGAGGAAGCT	1740
ACGAGAGGAG	GTTATCCTGA	CGATGGGACC	GTCCAGTGAT	GAAATAACCG	TGGCCCGGTT	1800
GAAGAGTTGC	CGTTACCTCA	AAGCAATCCT	AAACGAAACT	CTTCGACTAT	ACCCAAGTGT	1860
GCCTAGGAAC	GCGAGATTTG	CTACGAGGAA	TACGACGTT	CCTCGTGCGG	GAGGTCCAGA	1920
TGGATCGTTT	CCGATTTTGA	TAAGAAAGGG	CCAGCCAGTG	GGGTATTTCA	TTTGTGCTAC	1980
ACACTTGAAT	GAGAAGGTAT	ATGGGAATGA	TAGCCATGTG	TTTCGACCGG	AGAGATGGGC	2040
TGCGTTAGAG	GGCAAGAGTT	TGGGCTGGTC	GTATCTTCCA	TTCAACGGCG	GCCCAGAGAAG	2100
CTGCCTTGGT	CAGCAGTTTG	CAATCCTTGA	AGCTTCGTAT	GTTTGGGCTC	GATTGACACA	2160
GTGCTACACG	ACGATACAGC	TTAGAACTAC	CGAGTACCCA	CCAAAGAAAC	TCGTTTCATCT	2220
CACGATGAGT	CTTCTCAACG	GGGTGTACAT	CCGAACTAGA	ACTTGATTAT	GTGTTTATGG	2280
TTAATCGGGG	CAAAGCACTG	CAAGTCATTG	ATGTTTGTGT	AAGCCCAGCA	TTGGTGTTC	2340
GGAGTCATCA	TAACCAATGT	CTTGAAGGGT	TTGATTTTCT	TGACCTTCTT	CTTCCTGAGC	2400
TTCTTTCCGT	CAAACCTGTA	CAGAATGGCC	ATCATTTTCAG	GAACAACCAC	GTACGACGGC	2460
CGGTACCGCA	TCTGGAGTAT	CTCGCCGTCG	TTCAAGTAGC	ACGAAAACAG	CAACGACGTC	2520
ACCATCTGCT	TCCAATCTT	GACACCACCA	GATACCCCTG	CGGCTTCATG	GATCAAAAAC	2580
GTCCGGCAACC	CCGCGTATAT	GTCCATGTAA	TTCTCCATGG	CCACCTCCAT	CAACCACTG	2640
ATGGAGCGAC	TGACGGTGCC	ACCACTGCCC	TCGGTTGAGT	CAAGGCAGTA	TGATGCCGGG	2700
ATCCAGTACT	CCAATGGGAA	CCTCTGCACG	GTGTCGCTGC	AGTTTGTGAG	GCGTATTTTCG	2760
ATCCATGATC	GTTCTTTGGT	GCTGTAGTAT	AACGAGCTCT	TGGTGTCTCT	GAAATGGAAC	2820
AGGTTGGATG	TGTTGTTGAG	TTTGTCTGCG	TGCTTGGTTT	GCAAGTCTTC	GATCGAGCGT	2880
AGTGAGTAGA	CAGTTGGCGG	GGGTGGTGGC	TCGGGCTTTA	TTCTGTGTTT	GTGTTTCCTT	2940
CTTAGTCTTG	GAATGACGCT	GTTATCGACG	GTTCGTAAGT	TAAGTAGCGC	CAATATGAGA	3000
ATGTATATCC	GCATCACCCA	AGACTCTTCA	GCCTGTTACA	ACGACTGAGG	CTGTTGGCCG	3060
TGTGACCAAT	TGGTTTCTTT	GGTGACCTAG	ATTGGTCCCG	CAGGGAAAGC	AAGGGCTGCT	3120
AGGGGGGCAT	ACCAAACAAG	GTCTGTGAAT	CAGTATCTAT	GGTGCTACCA	TGTGTGTGGT	3180
TGGGGGGAAA	TTCCCGCATT	TTTGTGTAAC	GAAAGTTCTA	GAAAGTTCTC	GTGGGTTCTG	3240
AGAATCTGCT	GGAACCATCC	ACCCGCATT	CCGTTGCCAA	AGTGGGAAGA	GCAATCAACC	3300
CACCCTGCTT	TGCCCAATCA	GCCATTCCCC	TGGGAATATA	AATTCAAC		3348

(2) INFORMATION FOR SEQ ID NO:95:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 523 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

```

Met Ala Thr Gln Glu Ile Ile Asp Ser Val Leu Pro Tyr Leu Thr Lys
1           5           10           15
Trp Tyr Thr Val Ile Thr Ala Ala Val Leu Val Phe Leu Ile Ser Thr
20           25           30
Asn Ile Lys Asn Tyr Val Lys Ala Lys Lys Leu Lys Cys Val Asp Pro
35           40           45
Pro Tyr Leu Lys Asp Ala Gly Leu Thr Gly Ile Leu Ser Leu Ile Ala
50           55           60
Ala Ile Lys Ala Lys Asn Asp Gly Arg Leu Ala Asn Phe Ala Asp Glu
65           70           75           80
Val Phe Asp Glu Tyr Pro Asn His Thr Phe Tyr Leu Ser Val Ala Gly
85           90           95
Ala Leu Lys Ile Val Met Thr Val Asp Pro Glu Asn Ile Lys Ala Val
100          105          110
Leu Ala Thr Gln Phe Thr Asp Phe Ser Leu Gly Thr Arg His Ala His
115          120          125
Phe Ala Pro Leu Leu Gly Asp Gly Ile Phe Thr Leu Asp Gly Glu Gly
130          135          140
Trp Lys His Ser Arg Ala Met Leu Arg Pro Gln Phe Ala Arg Asp Gln
145          150          155          160
Ile Gly His Val Lys Ala Leu Glu Pro His Ile Gln Ile Met Ala Lys
165          170          175
Gln Ile Lys Leu Asn Gln Gly Lys Thr Phe Asp Ile Gln Glu Leu Phe
180          185          190
Phe Arg Phe Thr Val Asp Thr Ala Thr Glu Phe Leu Phe Gly Glu Ser
195          200          205
Val His Ser Leu Tyr Asp Glu Lys Leu Gly Ile Pro Thr Pro Asn Glu
210          215          220
Ile Pro Gly Arg Glu Asn Phe Ala Ala Ala Phe Asn Val Ser Gln His
225          230          235          240
Tyr Leu Ala Thr Arg Ser Tyr Ser Gln Thr Phe Tyr Phe Leu Thr Asn
245          250          255
Pro Lys Glu Phe Arg Asp Cys Asn Ala Lys Val His His Leu Ala Lys
260          265          270
Tyr Phe Val Asn Lys Ala Leu Asn Phe Thr Pro Glu Glu Leu Glu Glu
275          280          285
Lys Ser Lys Ser Gly Tyr Val Phe Leu Tyr Glu Leu Val Lys Gln Thr
290          295          300
Arg Asp Pro Lys Val Leu Gln Asp Gln Leu Leu Asn Ile Met Val Ala
305          310          315          320
Gly Arg Asp Thr Thr Ala Gly Leu Leu Ser Phe Ala Leu Phe Glu Leu
325          330          335
Ala Arg His Pro Glu Met Trp Ser Lys Leu Arg Glu Glu Ile Glu Val
340          345          350
Asn Phe Gly Val Gly Glu Asp Ser Arg Val Glu Glu Ile Thr Phe Glu
355          360          365
Ala Leu Lys Arg Cys Glu Tyr Leu Lys Ala Ile Leu Asn Glu Thr Leu
370          375          380
Arg Met Tyr Pro Ser Val Pro Val Asn Phe Arg Thr Ala Thr Arg Asp
385          390          395          400
Thr Thr Leu Pro Arg Gly Gly Gly Ala Asn Gly Thr Asp Pro Ile Tyr
405          410          415
Ile Pro Lys Gly Ser Thr Val Ala Tyr Val Val Tyr Lys Thr His Arg
420          425          430

```

```

Leu Glu Glu Tyr Tyr Gly Lys Asp Ala Asn Asp Phe Arg Pro Glu Arg
      435              440              445
Trp Phe Glu Pro Ser Thr Lys Lys Leu Gly Trp Ala Tyr Val Pro Phe
      450              455              460
Asn Gly Gly Pro Arg Val Cys Leu Gly Gln Gln Phe Ala Leu Thr Glu
465              470              475              480

Ala Ser Tyr Val Ile Thr Arg Leu Ala Gln Met Phe Glu Thr Val Ser
      485              490              495
Ser Asp Pro Gly Leu Glu Tyr Pro Pro Pro Lys Cys Ile His Leu Thr
      500              505              510
Met Ser His Asn Asp Gly Val Phe Val Lys Met
      515              520

```

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 522 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

```

Met Thr Val His Asp Ile Ile Ala Thr Tyr Phe Thr Lys Trp Tyr Val
1              5              10              15
Ile Val Pro Leu Ala Leu Ile Ala Tyr Arg Val Leu Asp Tyr Phe Tyr
      20              25              30
Gly Arg Tyr Leu Met Tyr Lys Leu Gly Ala Lys Pro Phe Phe Gln Lys
      35              40              45
Gln Thr Asp Gly Cys Phe Gly Phe Lys Ala Pro Leu Glu Leu Leu Lys
      50              55              60
Lys Lys Ser Asp Gly Thr Leu Ile Asp Phe Thr Leu Gln Arg Ile His
      65              70              75              80
Asp Leu Asp Arg Pro Asp Ile Pro Thr Phe Thr Phe Pro Val Phe Ser
      85              90              95
Ile Asn Leu Val Asn Thr Leu Glu Pro Glu Asn Ile Lys Ala Ile Leu
      100             105             110
Ala Thr Gln Phe Asn Asp Phe Ser Leu Gly Thr Arg His Ser His Phe
      115             120             125
Ala Pro Leu Leu Gly Asp Gly Ile Phe Thr Leu Asp Gly Ala Gly Trp
      130             135             140
Lys His Ser Arg Ser Met Leu Arg Pro Gln Phe Ala Arg Glu Gln Ile
      145             150             155             160
Ser His Val Lys Leu Leu Glu Pro His Val Gln Val Phe Phe Lys His
      165             170             175
Val Arg Lys Ala Gln Gly Lys Thr Phe Asp Ile Gln Glu Leu Phe Phe
      180             185             190
Arg Leu Thr Val Asp Ser Ala Thr Glu Phe Leu Phe Gly Glu Ser Val
      195             200             205
Glu Ser Leu Arg Asp Glu Ser Ile Gly Met Ser Ile Asn Ala Leu Asp
      210             215             220
Phe Asp Gly Lys Ala Gly Phe Ala Asp Ala Phe Asn Tyr Ser Gln Asn
      225             230             235             240
Tyr Leu Ala Ser Arg Ala Val Met Gln Gln Leu Tyr Trp Val Leu Asn
      245             250             255
Gly Lys Lys Phe Lys Glu Cys Asn Ala Lys Val His Lys Phe Ala Asp
      260             265             270

```



```

Tyr Tyr Val Asn Lys Ala Leu Asp Leu Thr Pro Glu Gln Leu Glu Lys
275 280 285
Gln Asp Gly Tyr Val Phe Leu Tyr Glu Leu Val Lys Gln Thr Arg Asp
290 295 300
Lys Gln Val Leu Arg Asp Gln Leu Leu Asn Ile Met Val Ala Gly Arg
305 310 315 320

Asp Thr Thr Ala Gly Leu Leu Ser Phe Val Phe Phe Glu Leu Ala Arg
325 330 335
Asn Pro Glu Val Thr Asn Lys Leu Arg Glu Glu Ile Glu Asp Lys Phe
340 345 350
Gly Leu Gly Glu Asn Ala Ser Val Glu Asp Ile Ser Phe Glu Ser Leu
355 360 365
Lys Ser Cys Glu Tyr Leu Lys Ala Val Leu Asn Glu Thr Leu Arg Leu
370 375 380
Tyr Pro Ser Val Pro Gln Asn Phe Arg Val Ala Thr Lys Asn Thr Thr
385 390 395 400
Leu Pro Arg Gly Gly Lys Asp Gly Leu Ser Pro Val Leu Val Arg
405 410 415
Lys Gly Gln Thr Val Ile Tyr Gly Val Tyr Ala Ala His Arg Asn Pro
420 425 430
Ala Val Tyr Gly Lys Asp Ala Leu Glu Phe Arg Pro Glu Arg Trp Phe
435 440 445
Glu Pro Glu Thr Lys Lys Leu Gly Trp Ala Phe Leu Pro Phe Asn Gly
450 455 460
Gly Pro Arg Ile Cys Leu Gly Gln Gln Phe Ala Leu Thr Glu Ala Ser
465 470 475 480
Tyr Val Thr Val Arg Leu Leu Gln Glu Phe Ala His Leu Ser Met Asp
485 490 495
Pro Asp Thr Glu Tyr Pro Pro Lys Lys Met Ser His Leu Thr Met Ser
500 505 510
Leu Phe Asp Gly Ala Asn Ile Glu Met Tyr
515 520

```

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 522 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

```

Met Thr Ala Gln Asp Ile Ile Ala Thr Tyr Ile Thr Lys Trp Tyr Val
1 5 10 15
Ile Val Pro Leu Ala Leu Ile Ala Tyr Arg Val Leu Asp Tyr Phe Tyr
20 25 30
Gly Arg Tyr Leu Met Tyr Lys Leu Gly Ala Lys Pro Phe Phe Gln Lys
35 40 45
Gln Thr Asp Gly Tyr Phe Gly Phe Lys Ala Pro Leu Glu Leu Leu Lys
50 55 60
Lys Lys Ser Asp Gly Thr Leu Ile Asp Phe Thr Leu Glu Arg Ile Gln
65 70 75 80
Ala Leu Asn Arg Pro Asp Ile Pro Thr Phe Thr Phe Pro Ile Phe Ser
85 90 95
Ile Asn Leu Ile Ser Thr Leu Glu Pro Glu Asn Ile Lys Ala Ile Leu
100 105 110

```

```

Ala Thr Gln Phe Asn Asp Phe Ser Leu Gly Thr Arg His Ser His Phe
      115                      120                      125
Ala Pro L u Leu Gly Asp Gly Ile Phe Thr Leu Asp Gly Ala Gly Trp
      130                      135                      140
Lys His Ser Arg Ser Met Leu Arg Pro Gln Phe Ala Arg Glu Gln Ile
      145                      150                      155                      160
Ser His Val Lys Leu Leu Glu Pro His Met Gln Val Phe Phe Lys His
      165                      170                      175
Val Arg Lys Ala Gln Gly Lys Thr Phe Asp Ile Gln Glu Leu Phe Phe
      180                      185                      190
Arg Leu Thr Val Asp Ser Ala Thr Glu Phe Leu Phe Gly Glu Ser Val
      195                      200                      205
Glu Ser Leu Arg Asp Glu Ser Ile Gly Met Ser Ile Asn Ala Leu Asp
      210                      215                      220
Phe Asp Gly Lys Ala Gly Phe Ala Asp Ala Phe Asn Tyr Ser Gln Asn
      225                      230                      235                      240
Tyr Leu Ala Ser Arg Ala Val Met Gln Gln Leu Tyr Trp Val Leu Asn
      245                      250                      255
Gly Lys Lys Phe Lys Glu Cys Asn Ala Lys Val His Lys Phe Ala Asp
      260                      265                      270
Tyr Tyr Val Ser Lys Ala Leu Asp Leu Thr Pro Glu Gln Leu Glu Lys
      275                      280                      285
Gln Asp Gly Tyr Val Phe Leu Tyr Glu Leu Val Lys Gln Thr Arg Asp
      290                      295                      300
Arg Gln Val Leu Arg Asp Gln Leu Leu Asn Ile Met Val Ala Gly Arg
      305                      310                      315                      320
Asp Thr Thr Ala Gly Leu Leu Ser Phe Val Phe Phe Glu Leu Ala Arg
      325                      330                      335
Asn Pro Glu Val Thr Asn Lys Leu Arg Glu Glu Ile Glu Asp Lys Phe
      340                      345                      350
Gly Leu Gly Glu Asn Ala Arg Val Glu Asp Ile Ser Phe Glu Ser Leu
      355                      360                      365
Lys Ser Cys Glu Tyr Leu Lys Ala Val Leu Asn Glu Thr Leu Arg Leu
      370                      375                      380
Tyr Pro Ser Val Pro Gln Asn Phe Arg Val Ala Thr Lys Asn Thr Thr
      385                      390                      395                      400
Leu Pro Arg Gly Gly Gly Lys Asp Gly Leu Ser Pro Val Leu Val Arg
      405                      410                      415
Lys Gly Gln Thr Val Met Tyr Gly Val Tyr Ala Ala His Arg Asn Pro
      420                      425                      430
Ala Val Tyr Gly Lys Asp Ala Leu Glu Phe Arg Pro Glu Arg Trp Phe
      435                      440                      445
Glu Pro Glu Thr Lys Lys Leu Gly Trp Ala Phe Leu Pro Phe Asn Gly
      450                      455                      460
Gly Pro Arg Ile Cys Leu Gly Gln Gln Phe Ala Leu Thr Glu Ala Ser
      465                      470                      475                      480
Tyr Val Thr Val Arg Leu Leu Gln Glu Phe Gly His Leu Ser Met Asp
      485                      490                      495
Pro Asn Thr Glu Tyr Pro Pro Arg Lys Met Ser His Leu Thr Met Ser
      500                      505                      510
Leu Phe Asp Gly Ala Asn Ile Glu Met Tyr
      515                      520

```

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 540 amino acids

(B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

Met	Ser	Ser	Ser	Pro	Ser	Phe	Ala	Gln	Glu	Val	Leu	Ala	Thr	Thr	Ser
1				5					10					15	
Pro	Tyr	Ile	Glu	Tyr	Phe	Leu	Asp	Asn	Tyr	Thr	Arg	Trp	Tyr	Tyr	Phe
			20					25					30		
Ile	Pro	Leu	Val	Leu	Leu	Ser	Leu	Asn	Phe	Ile	Ser	Leu	Leu	His	Thr
		35					40				45				
Arg	Tyr	Leu	Glu	Arg	Arg	Phe	His	Ala	Lys	Pro	Leu	Gly	Asn	Phe	Val
	50					55					60				
Arg	Asp	Pro	Thr	Phe	Gly	Ile	Ala	Thr	Pro	Leu	Leu	Leu	Ile	Tyr	Leu
65					70					75				80	
Lys	Ser	Lys	Gly	Thr	Val	Met	Lys	Phe	Ala	Trp	Gly	Leu	Trp	Asn	Asn
			85						90					95	
Lys	Tyr	Ile	Val	Arg	Asp	Pro	Lys	Tyr	Lys	Thr	Thr	Gly	Leu	Arg	Ile
			100					105					110		
Val	Gly	Leu	Pro	Leu	Ile	Glu	Thr	Met	Asp	Pro	Glu	Asn	Ile	Lys	Ala
		115					120				125				
Val	Leu	Ala	Thr	Gln	Phe	Asn	Asp	Phe	Ser	Leu	Gly	Thr	Arg	His	Asp
	130					135					140				
Phe	Leu	Tyr	Ser	Leu	Leu	Gly	Asp	Gly	Ile	Phe	Thr	Leu	Asp	Gly	Ala
145					150					155				160	
Gly	Trp	Lys	His	Ser	Arg	Thr	Met	Leu	Arg	Pro	Gln	Phe	Ala	Arg	Glu
			165						170					175	
Gln	Val	Ser	His	Val	Lys	Leu	Leu	Glu	Pro	His	Val	Gln	Val	Phe	Phe
			180					185					190		
Lys	His	Val	Arg	Lys	His	Arg	Gly	Gln	Thr	Phe	Asp	Ile	Gln	Glu	Leu
	195						200					205			
Phe	Phe	Arg	Leu	Thr	Val	Asp	Ser	Ala	Thr	Glu	Phe	Leu	Phe	Gly	Glu
	210					215					220				
Ser	Ala	Glu	Ser	Leu	Arg	Asp	Glu	Ser	Ile	Gly	Leu	Thr	Pro	Thr	Thr
225				230						235				240	
Lys	Asp	Phe	Asp	Gly	Arg	Arg	Asp	Phe	Ala	Asp	Ala	Phe	Asn	Tyr	Ser
			245					250						255	
Gln	Thr	Tyr	Gln	Ala	Tyr	Arg	Phe	Leu	Leu	Gln	Gln	Met	Tyr	Trp	Ile
			260					265					270		
Leu	Asn	Gly	Ser	Glu	Phe	Arg	Lys	Ser	Ile	Ala	Val	Val	His	Lys	Phe
	275						280					285			
Ala	Asp	His	Tyr	Val	Gln	Lys	Ala	Leu	Glu	Leu	Thr	Asp	Asp	Asp	Leu
	290					295					300				
Gln	Lys	Gln	Asp	Gly	Tyr	Val	Phe	Leu	Tyr	Glu	Leu	Ala	Lys	Gln	Thr
305				310						315				320	
Arg	Asp	Pro	Lys	Val	Leu	Arg	Asp	Gln	Leu	Leu	Asn	Ile	Leu	Val	Ala
			325					330						335	
Gly	Arg	Asp	Thr	Thr	Ala	Gly	Leu	Leu	Ser	Phe	Val	Phe	Tyr	Glu	Leu
		340					345						350		
Ser	Arg	Asn	Pro	Glu	Val	Phe	Ala	Lys	Leu	Arg	Glu	Glu	Val	Glu	Asn
		355					360					365			
Arg	Phe	Gly	Leu	Gly	Glu	Glu	Ala	Arg	Val	Glu	Glu	Ile	Ser	Phe	Glu
	370					375					380				
Ser	Leu	Lys	Ser	Cys	Glu	Tyr	Leu	Lys	Ala	Val	Ile	Asn	Glu	Thr	Leu
385					390					395					400

```

Arg Leu Tyr Pro Ser Val Pro His Asn Phe Arg Val Ala Thr Arg Asn
      405                      410                      415
Thr Thr Leu Pro Arg Gly Gly Gly Glu Asp Gly Tyr Ser Pr Ile Val
      420                      425                      430
Val Lys Lys Gly Gln Val Val Met Tyr Thr Val Ile Ala Thr His Arg
      435                      440                      445
Asp Pro Ser Ile Tyr Gly Ala Asp Ala Asp Val Phe Arg Pro Glu Arg
      450                      455                      460
Trp Phe Glu Pro Glu Thr Arg Lys Leu Gly Trp Ala Tyr Val Pro Phe
      465                      470                      475                      480
Asn Gly Gly Pro Arg Ile Cys Leu Gly Gln Gln Phe Ala Leu Thr Glu
      485                      490                      495
Ala Ser Tyr Val Thr Val Arg Leu Leu Gln Glu Phe Ala His Leu Ser
      500                      505                      510
Met Asp Pro Asp Thr Glu Tyr Pro Pro Lys Leu Gln Asn Thr Leu Thr
      515                      520                      525
Leu Ser Leu Phe Asp Gly Ala Asp Val Arg Met Tyr
      530                      535                      540

```

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 540 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

```

Met Ser Ser Ser Pro Ser Phe Ala Gln Glu Val Leu Ala Thr Thr Ser
1          5          10          15
Pro Tyr Ile Glu Tyr Phe Leu Asp Asn Tyr Thr Arg Trp Tyr Tyr Phe
      20          25          30
Ile Pro Leu Val Leu Leu Ser Leu Asn Phe Ile Ser Leu Leu His Thr
      35          40          45
Lys Tyr Leu Glu Arg Arg Phe His Ala Lys Pro Leu Gly Asn Val Val
      50          55          60
Leu Asp Pro Thr Phe Gly Ile Ala Thr Pro Leu Ile Leu Ile Tyr Leu
      65          70          75          80
Lys Ser Lys Gly Thr Val Met Lys Phe Ala Trp Ser Phe Trp Asn Asn
      85          90          95
Lys Tyr Ile Val Lys Asp Pro Lys Tyr Lys Thr Thr Gly Leu Arg Ile
      100         105         110
Val Gly Leu Pro Leu Ile Glu Thr Ile Asp Pro Glu Asn Ile Lys Ala
      115         120         125
Val Leu Ala Thr Gln Phe Asn Asp Phe Ser Leu Gly Thr Arg His Asp
      130         135         140
Phe Leu Tyr Ser Leu Leu Gly Asp Gly Ile Phe Thr Leu Asp Gly Ala
      145         150         155         160
Gly Trp Lys His Ser Arg Thr Met Leu Arg Pro Gln Phe Ala Arg Glu
      165         170         175
Gln Val Ser His Val Lys Leu Leu Glu Pro His Val Gln Val Phe Phe
      180         185         190
Lys His Val Arg Lys His Arg Gly Gln Thr Phe Asp Ile Gln Glu Leu
      195         200         205
Phe Phe Arg Leu Thr Val Asp Ser Ala Thr Glu Phe Leu Phe Gly Glu
      210         215         220

```

```

Ser Ala Glu Ser Leu Arg Asp Asp Ser Val Gly Leu Thr Pro Thr Thr
225                230                235                240
Lys Asp Phe Glu Gly Arg Gly Asp Phe Ala Asp Ala Phe Asn Tyr Ser
                245                250                255
Gln Thr Tyr Gln Ala Tyr Arg Phe Leu Leu Gln Gln Met Tyr Trp Ile
                260                265                270
Leu Asn Gly Ala Glu Phe Arg Lys Ser Ile Ala Ile Val His Lys Phe
                275                280                285
Ala Asp His Tyr Val Gln Lys Ala Leu Glu Leu Thr Asp Asp Asp Leu
                290                295                300
Gln Lys Gln Asp Gly Tyr Val Phe Leu Tyr Glu Leu Ala Lys Gln Thr
305                310                315                320
Arg Asp Pro Lys Val Leu Arg Asp Gln Leu Leu Asn Ile Leu Val Ala
                325                330                335
Gly Arg Asp Thr Thr Ala Gly Leu Leu Ser Phe Val Phe Tyr Glu Leu
                340                345                350
Ser Arg Asn Pro Glu Val Phe Ala Lys Leu Arg Glu Glu Val Glu Asn
                355                360                365
Arg Phe Gly Leu Gly Glu Glu Ala Arg Val Glu Glu Ile Ser Phe Glu
370                375                380
Ser Leu Lys Ser Cys Glu Tyr Leu Lys Ala Val Ile Asn Glu Ala Leu
385                390                395                400
Arg Leu Tyr Pro Ser Val Pro His Asn Phe Arg Val Ala Thr Arg Asn
                405                410                415
Thr Thr Leu Pro Arg Gly Gly Gly Lys Asp Gly Cys Ser Pro Ile Val
                420                425                430
Val Lys Lys Gly Gln Val Val Met Tyr Thr Val Ile Gly Thr His Arg
435                440                445
Asp Pro Ser Ile Tyr Gly Ala Asp Ala Asp Val Phe Arg Pro Glu Arg
450                455                460
Trp Phe Glu Pro Glu Thr Arg Lys Leu Gly Trp Ala Tyr Val Pro Phe
465                470                475                480
Asn Gly Gly Pro Arg Ile Cys Leu Gly Gln Gln Phe Ala Leu Thr Glu
                485                490                495
Ala Ser Tyr Val Thr Val Arg Leu Leu Gln Glu Phe Gly Asn Leu Ser
                500                505                510
Leu Asp Pro Asn Ala Glu Tyr Pro Lys Leu Gln Asn Thr Leu Thr
515                520                525
Leu Ser Leu Phe Asp Gly Ala Asp Val Arg Met Phe
530                535                540

```

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 517 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

```

Met Ile Glu Gln Leu Leu Glu Tyr Trp Tyr Val Val Val Pro Val Leu
1           5           10           15
Tyr Ile Ile Lys Gln Leu Leu Ala Tyr Thr Lys Thr Arg Val Leu Met
20           25           30
Lys Lys Leu Gly Ala Ala Pro Val Thr Asn Lys Leu Tyr Asp Asn Ala
35           40           45

```

Phe Gly Ile Val Asn Gly Trp Lys Ala Leu Gln Phe Lys Lys Glu Gly
 50 55 60
 Arg Ala Gln Glu Tyr Asn Asp Tyr Lys Phe Asp His Ser Lys Asn Pro
 65 70 75 80
 Ser Val Gly Thr Tyr Val Ser Ile Leu Phe Gly Thr Arg Ile Val Val
 85 90 95
 Thr Lys Asp Pro Glu Asn Ile Lys Ala Ile Leu Ala Thr Gln Phe Gly
 100 105 110
 Asp Phe Ser Leu Gly Lys Arg His Thr Leu Phe Lys Pro Leu Leu Gly
 115 120 125
 Asp Gly Ile Phe Thr Leu Asp Gly Glu Gly Trp Lys His Ser Arg Ala
 130 135 140
 Met Leu Arg Pro Gln Phe Ala Arg Glu Gln Val Ala His Val Thr Ser
 145 150 155 160
 Leu Glu Pro His Phe Gln Leu Leu Lys Lys His Ile Leu Lys His Lys
 165 170 175
 Gly Glu Tyr Phe Asp Ile Gln Glu Leu Phe Phe Arg Phe Thr Val Asp
 180 185 190
 Ser Ala Thr Glu Phe Leu Phe Gly Glu Ser Val His Ser Leu Lys Asp
 195 200 205
 Glu Ser Ile Gly Ile Asn Gln Asp Asp Ile Asp Phe Ala Gly Arg Lys
 210 215 220
 Asp Phe Ala Glu Ser Phe Asn Lys Ala Gln Glu Tyr Leu Ala Ile Arg
 225 230 235 240
 Thr Leu Val Gln Thr Phe Tyr Trp Leu Val Asn Asn Lys Glu Phe Arg
 245 250 255
 Asp Cys Thr Lys Leu Val His Lys Phe Thr Asn Tyr Tyr Val Gln Lys
 260 265 270
 Ala Leu Asp Ala Ser Pro Glu Glu Leu Glu Lys Gln Ser Gly Tyr Val
 275 280 285
 Phe Leu Tyr Glu Leu Val Lys Gln Thr Arg Asp Pro Asn Val Leu Arg
 290 295 300
 Asp Gln Ser Leu Asn Ile Leu Leu Ala Gly Arg Asp Thr Thr Ala Gly
 305 310 315 320
 Leu Leu Ser Phe Ala Val Phe Glu Leu Ala Arg His Pro Glu Ile Trp
 325 330 335
 Ala Lys Leu Arg Glu Glu Ile Glu Gln Gln Phe Gly Leu Gly Glu Asp
 340 345 350
 Ser Arg Val Glu Glu Ile Thr Phe Glu Ser Leu Lys Arg Cys Glu Tyr
 355 360 365
 Leu Lys Ala Phe Leu Asn Glu Thr Leu Arg Ile Tyr Pro Ser Val Pro
 370 375 380
 Arg Asn Phe Arg Ile Ala Thr Lys Asn Thr Thr Leu Pro Arg Gly Gly
 385 390 395 400
 Gly Ser Asp Gly Thr Ser Pro Ile Leu Ile Gln Lys Gly Glu Ala Val
 405 410 415
 Ser Tyr Gly Ile Asn Ser Thr His Leu Asp Pro Val Tyr Tyr Gly Pro
 420 425 430
 Asp Ala Ala Glu Phe Arg Pro Glu Arg Trp Phe Glu Pro Ser Thr Lys
 435 440 445
 Lys Leu Gly Trp Ala Tyr Leu Pro Phe Asn Gly Gly Pro Arg Ile Cys
 450 455 460
 Leu Gly Gln Gln Phe Ala Leu Thr Glu Ala Gly Tyr Val Leu Val Arg
 465 470 475 480
 Leu Val Gln Glu Phe Ser His Val Arg Leu Asp Pro Asp Glu Val Tyr
 485 490 495

Pro Pro Lys Arg Leu Thr Asn Leu Thr Met Cys Leu Gln Asp Gly Ala
 500 505 510
 Ile Val Lys Phe Asp
 515

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 517 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

Met Ile Glu Gln Ile Leu Glu Tyr Trp Tyr Ile Val Val Pro Val Leu
 1 5 10 15
 Tyr Ile Ile Lys Gln Leu Ile Ala Tyr Ser Lys Thr Arg Val Leu Met
 20 25 30
 Lys Gln Leu Gly Ala Ala Pro Ile Thr Asn Gln Leu Tyr Asp Asn Val
 35 40 45
 Phe Gly Ile Val Asn Gly Trp Lys Ala Leu Gln Phe Lys Lys Glu Gly
 50 55 60
 Arg Ala Gln Glu Tyr Asn Asp His Lys Phe Asp Ser Ser Lys Asn Pro
 65 70 75 80
 Ser Val Gly Thr Tyr Val Ser Ile Leu Phe Gly Thr Lys Ile Val Val
 85 90 95
 Thr Lys Asp Pro Glu Asn Ile Lys Ala Ile Leu Ala Thr Gln Phe Gly
 100 105 110
 Asp Phe Ser Leu Gly Lys Arg His Ala Leu Phe Lys Pro Leu Leu Gly
 115 120 125
 Asp Gly Ile Phe Thr Leu Asp Gly Glu Gly Trp Lys His Ser Arg Ser
 130 135 140
 Met Leu Arg Pro Gln Phe Ala Arg Glu Gln Val Ala His Val Thr Ser
 145 150 155 160
 Leu Glu Pro His Phe Gln Leu Leu Lys Lys His Ile Leu Lys His Lys
 165 170 175
 Gly Glu Tyr Phe Asp Ile Gln Glu Leu Phe Phe Arg Phe Thr Val Asp
 180 185 190
 Ser Ala Thr Glu Phe Leu Phe Gly Glu Ser Val His Ser Leu Lys Asp
 195 200 205
 Glu Thr Ile Gly Ile Asn Gln Asp Asp Ile Asp Phe Ala Gly Arg Lys
 210 215 220
 Asp Phe Ala Glu Ser Phe Asn Lys Ala Gln Glu Tyr Leu Ser Ile Arg
 225 230 235 240
 Ile Leu Val Gln Thr Phe Tyr Trp Leu Ile Asn Asn Lys Glu Phe Arg
 245 250 255
 Asp Cys Thr Lys Leu Val His Lys Phe Thr Asn Tyr Tyr Val Gln Lys
 260 265 270
 Ala Leu Asp Ala Thr Pro Glu Glu Leu Glu Lys Gln Gly Gly Tyr Val
 275 280 285
 Phe Leu Tyr Glu Leu Val Lys Gln Thr Arg Asp Pro Lys Val Leu Arg
 290 295 300
 Asp Gln Ser Leu Asn Ile Leu Leu Ala Gly Arg Asp Thr Thr Ala Gly
 305 310 315 320
 Leu Leu Ser Phe Ala Val Phe Glu Leu Ala Arg Asn Pro His Ile Trp
 325 330 335

```

Ala Lys Leu Arg Glu Glu Il  Glu Gln Gln Phe Gly Leu Gly Glu Asp
      340      345      350
Ser Arg Val Glu Glu Ile Thr Phe Glu Ser Leu Lys Arg Cys Glu Tyr
      355      360      365
Leu Lys Ala Phe Leu Asn Glu Thr Leu Arg Val Tyr Pro Ser Val Pro
      370      375      380
Arg Asn Phe Arg Ile Ala Thr Lys Asn Thr Thr Leu Pro Arg Gly Gly
      385      390      395      400
Gly Pro Asp Gly Thr Gln Pro Ile Leu Ile Gln Lys Gly Glu Gly Val
      405      410      415
Ser Tyr Gly Ile Asn Ser Thr His Leu Asp Pro Val Tyr Tyr Gly Pro
      420      425      430
Asp Ala Ala Glu Phe Arg Pro Glu Arg Trp Phe Glu Pro Ser Thr Arg
      435      440      445
Lys Leu Gly Trp Ala Tyr Leu Pro Phe Asn Gly Gly Pro Arg Ile Cys
      450      455      460
Leu Gly Gln Gln Phe Ala Leu Thr Glu Ala Gly Tyr Val Leu Val Arg
      465      470      475      480
Leu Val Gln Glu Phe Ser His Ile Arg Leu Asp Pro Asp Glu Val Tyr
      485      490      495
Pro Pro Lys Arg Leu Thr Asn Leu Thr Met Cys Leu Gln Asp Gly Ala
      500      505      510
Ile Val Lys Phe Asp
      515

```

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 512 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

```

Met Leu Asp Gln Ile Leu His Tyr Trp Tyr Ile Val Leu Pro Leu Leu
1      5      10      15
Ala Ile Ile Asn Gln Ile Val Ala His Val Arg Thr Asn Tyr Leu Met
20      25      30
Lys Lys Leu Gly Ala Lys Pro Phe Thr His Val Gln Arg Asp Gly Trp
35      40      45
Leu Gly Phe Lys Phe Gly Arg Glu Phe Leu Lys Ala Lys Ser Ala Gly
50      55      60
Arg Leu Val Asp Leu Ile Ile Ser Arg Phe His Asp Asn Glu Asp Thr
65      70      75      80
Phe Ser Ser Tyr Ala Phe Gly Asn His Val Val Phe Thr Arg Asp Pro
85      90      95
Glu Asn Ile Lys Ala Leu Leu Ala Thr Gln Phe Gly Asp Phe Ser Leu
100      105      110
Gly Ser Arg Val Lys Phe Phe Lys Pro Leu Leu Gly Tyr Gly Ile Phe
115      120      125
Thr Leu Asp Ala Glu Gly Trp Lys His Ser Arg Ala Met Leu Arg Pro
130      135      140
Gln Phe Ala Arg Glu Gln Val Ala His Val Thr Ser Leu Glu Pro His
145      150      155      160
Phe Gln Leu Leu Lys Lys His Ile Leu Lys His Lys Gly Glu Tyr Phe
165      170      175

```



```

Asp Ile Gln Glu Leu Phe Phe Arg Phe Thr Val Asp Ser Ala Thr Glu
      180              185              190
Phe Leu Phe Gly Glu Ser Val His Ser Leu Lys Asp Glu Glu Ile Gly
      195              200              205
Tyr Asp Thr Lys Asp Met Ser Glu Glu Arg Arg Arg Phe Ala Asp Ala
      210              215              220

Phe Asn Lys Ser Gln Val Tyr Val Ala Thr Arg Val Ala Leu Gln Asn
225              230              235              240
Leu Tyr Trp Leu Val Asn Asn Lys Glu Phe Lys Glu Cys Asn Asp Ile
      245              250              255
Val His Lys Phe Thr Asn Tyr Tyr Val Gln Lys Ala Leu Asp Ala Thr
      260              265              270
Pro Glu Glu Leu Glu Lys Gln Gly Gly Tyr Val Phe Leu Tyr Glu Leu
      275              280              285
Val Lys Gln Thr Arg Asp Pro Lys Val Leu Arg Asp Gln Ser Leu Asn
      290              295              300
Ile Leu Leu Ala Gly Arg Asp Thr Thr Ala Gly Leu Leu Ser Phe Ala
305              310              315              320
Val Phe Glu Leu Ala Arg Asn Pro His Ile Trp Ala Lys Leu Arg Glu
      325              330              335
Glu Ile Glu Gln Gln Phe Gly Leu Gly Glu Asp Ser Arg Val Glu Glu
      340              345              350
Ile Thr Phe Glu Ser Leu Lys Arg Cys Glu Tyr Leu Lys Ala Val Leu
      355              360              365
Asn Glu Thr Leu Arg Leu His Pro Ser Val Pro Arg Asn Ala Arg Phe
      370              375              380
Ala Ile Lys Asp Thr Thr Leu Pro Arg Gly Gly Gly Pro Asn Gly Lys
385              390              395              400
Asp Pro Ile Leu Ile Arg Lys Asp Glu Val Val Gln Tyr Ser Ile Ser
      405              410              415
Ala Thr Gln Thr Asn Pro Ala Tyr Tyr Gly Ala Asp Ala Ala Asp Phe
      420              425              430
Arg Pro Glu Arg Trp Phe Glu Pro Ser Thr Arg Asn Leu Gly Trp Ala
      435              440              445
Phe Leu Pro Phe Asn Gly Gly Pro Arg Ile Cys Leu Gly Gln Gln Phe
      450              455              460
Ala Leu Thr Glu Ala Gly Tyr Val Leu Val Arg Leu Val Gln Glu Phe
465              470              475              480
Pro Asn Leu Ser Gln Asp Pro Glu Thr Lys Tyr Pro Pro Pro Arg Leu
      485              490              495
Ala His Leu Thr Met Cys Leu Phe Asp Gly Ala His Val Lys Met Ser
      500              505              510

```

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 512 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

```

Met Leu Asp Gln Ile Phe His Tyr Trp Tyr Ile Val Leu Pro Leu Leu
 1           5           10           15
Val Ile Ile Lys Gln Ile Val Ala His Ala Arg Thr Asn Tyr Leu Met
20           25           30

```

Lys Lys Leu Gly Ala Lys Pro Phe Thr His Val Gln Leu Asp Gly Trp
 35 40 45
 Phe Gly Phe Lys Phe Gly Arg Glu Phe Leu Lys Ala Lys Ser Ala Gly
 50 55 60
 Arg Gln Val Asp Leu Ile Ile Ser Arg Phe His Asp Asn Glu Asp Thr
 65 70 75 80
 Phe Ser Ser Tyr Ala Phe Gly Asn His Val Val Phe Thr Arg Asp Pro
 85 90 95
 Glu Asn Ile Lys Ala Leu Leu Ala Thr Gln Phe Gly Asp Phe Ser Leu
 100 105 110
 Gly Ser Arg Val Lys Phe Phe Lys Pro Leu Leu Gly Tyr Gly Ile Phe
 115 120 125
 Thr Leu Asp Gly Glu Gly Trp Lys His Ser Arg Ala Met Leu Arg Pro
 130 135 140
 Gln Phe Ala Arg Glu Gln Val Ala His Val Thr Ser Leu Glu Pro His
 145 150 155 160
 Phe Gln Leu Leu Lys Lys His Ile Leu Lys His Lys Gly Glu Tyr Phe
 165 170 175
 Asp Ile Gln Glu Leu Phe Phe Arg Phe Thr Val Asp Ser Ala Thr Glu
 180 185 190
 Phe Leu Phe Gly Glu Ser Val His Ser Leu Arg Asp Glu Glu Ile Gly
 195 200 205
 Tyr Asp Thr Lys Asp Met Ala Glu Glu Arg Arg Lys Phe Ala Asp Ala
 210 215 220
 Phe Asn Lys Ser Gln Val Tyr Leu Ser Thr Arg Val Ala Leu Gln Thr
 225 230 235 240
 Leu Tyr Trp Leu Val Asn Asn Lys Glu Phe Lys Glu Cys Asn Asp Ile
 245 250 255
 Val His Lys Phe Thr Asn Tyr Tyr Val Gln Lys Ala Leu Asp Ala Thr
 260 265 270
 Pro Glu Glu Leu Glu Lys Gln Gly Gly Tyr Val Phe Leu Tyr Glu Leu
 275 280 285
 Ala Lys Gln Thr Lys Asp Pro Asn Val Leu Arg Asp Gln Ser Leu Asn
 290 295 300
 Ile Leu Leu Ala Gly Arg Asp Thr Thr Ala Gly Leu Leu Ser Phe Ala
 305 310 315 320
 Val Phe Glu Leu Ala Arg Asn Pro His Ile Trp Ala Lys Leu Arg Glu
 325 330 335
 Glu Ile Glu Ser His Phe Gly Leu Gly Glu Asp Ser Arg Val Glu Glu
 340 345 350
 Ile Thr Phe Glu Ser Leu Lys Arg Cys Glu Tyr Leu Lys Ala Val Leu
 355 360 365
 Asn Glu Thr Leu Arg Leu His Pro Ser Val Pro Arg Asn Ala Arg Phe
 370 375 380
 Ala Ile Lys Asp Thr Thr Leu Pro Arg Gly Gly Gly Pro Asn Gly Lys
 385 390 395 400
 Asp Pro Ile Leu Ile Arg Lys Asn Glu Val Val Gln Tyr Ser Ile Ser
 405 410 415
 Ala Thr Gln Thr Asn Pro Ala Tyr Tyr Gly Ala Asp Ala Ala Asp Phe
 420 425 430
 Arg Pro Glu Arg Trp Phe Glu Pro Ser Thr Arg Asn Leu Gly Trp Ala
 435 440 445
 Tyr Leu Pro Phe Asn Gly Gly Pro Arg Ile Cys Leu Gly Gln Gln Phe
 450 455 460
 Ala Leu Thr Glu Ala Gly Tyr Val Leu Val Arg Leu Val Gln Glu Phe
 465 470 475 480

Pro	Ser	Leu	Ser	Gln	Asp	Pro	Glu	Thr	Glu	Tyr	Pro	Pro	Pro	Arg	Leu
				485					490					495	
Ala	His	Leu	Thr	Met	Cys	Leu	Phe	Asp	Gly	Ala	Tyr	Val	Lys	Met	Gln
				500				505					510		

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 499 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

Met	Ala	Ile	Ser	Ser	Leu	Leu	Ser	Trp	Asp	Val	Ile	Cys	Val	Val	Phe
1			5					10						15	
Ile	Cys	Val	Cys	Val	Tyr	Phe	Gly	Tyr	Glu	Tyr	Cys	Tyr	Thr	Lys	Tyr
		20					25						30		
Leu	Met	His	Lys	His	Gly	Ala	Arg	Glu	Ile	Glu	Asn	Val	Ile	Asn	Asp
		35					40					45			
Gly	Phe	Phe	Gly	Phe	Arg	Leu	Pro	Leu	Leu	Leu	Met	Arg	Ala	Ser	Asn
	50					55					60				
Glu	Gly	Arg	Leu	Ile	Glu	Phe	Ser	Val	Lys	Arg	Phe	Glu	Ser	Ala	Pro
65				70					75					80	
His	Pro	Gln	Asn	Lys	Thr	Leu	Val	Asn	Arg	Ala	Leu	Ser	Val	Pro	Val
			85					90						95	
Ile	Leu	Thr	Lys	Asp	Pro	Val	Asn	Ile	Lys	Ala	Met	Leu	Ser	Thr	Gln
		100						105					110		
Phe	Asp	Asp	Phe	Ser	Leu	Gly	Leu	Arg	Leu	His	Gln	Phe	Ala	Pro	Leu
	115					120					125				
Leu	Gly	Lys	Gly	Ile	Phe	Thr	Leu	Asp	Gly	Pro	Glu	Trp	Lys	Gln	Ser
	130					135					140				
Arg	Ser	Met	Leu	Arg	Pro	Gln	Phe	Ala	Lys	Asp	Arg	Val	Ser	His	Ile
145					150				155					160	
Leu	Asp	Leu	Glu	Pro	His	Phe	Val	Leu	Leu	Arg	Lys	His	Ile	Asp	Gly
			165					170						175	
His	Asn	Gly	Asp	Tyr	Phe	Asp	Ile	Gln	Glu	Leu	Tyr	Phe	Arg	Phe	Ser
	180						185						190		
Met	Asp	Val	Ala	Thr	Gly	Phe	Leu	Phe	Gly	Glu	Ser	Val	Gly	Ser	Leu
	195						200					205			
Lys	Asp	Glu	Asp	Ala	Arg	Phe	Leu	Glu	Ala	Phe	Asn	Glu	Ser	Gln	Lys
	210					215					220				
Tyr	Leu	Ala	Thr	Arg	Ala	Thr	Leu	His	Glu	Leu	Tyr	Phe	Leu	Cys	Asp
225					230				235					240	
Gly	Phe	Arg	Phe	Arg	Gln	Tyr	Asn	Lys	Val	Val	Arg	Lys	Phe	Cys	Ser
			245					250						255	
Gln	Cys	Val	His	Lys	Ala	Leu	Asp	Val	Ala	Pro	Glu	Asp	Thr	Ser	Glu
		260					265						270		
Tyr	Val	Phe	Leu	Arg	Glu	Leu	Val	Lys	His	Thr	Arg	Asp	Pro	Val	Val
	275					280					285				
Leu	Gln	Asp	Gln	Ala	Leu	Asn	Val	Leu	Leu	Ala	Gly	Arg	Asp	Thr	Thr
	290					295					300				
Ala	Ser	Leu	Leu	Ser	Phe	Ala	Thr	Phe	Glu	Leu	Ala	Arg	Asn	Asp	His
305					310				315					320	
Met	Trp	Arg	Lys	Leu	Arg	Glu	Glu	Val	Ile	Leu	Thr	Met	Gly	Pro	Ser
			325					330					335		

Ser Asp Glu Ile Thr Val Ala Gly Leu Lys Ser Cys Arg Tyr Leu Lys
 340 345 350
 Ala Ile Leu Asn Glu Thr Leu Arg Leu Tyr Pro Ser Val Pro Arg Asn
 355 360 365
 Ala Arg Phe Ala Thr Arg Asn Thr Thr Leu Pro Arg Gly Gly Gly Pro
 370 375 380
 Asp Gly Ser Phe Pro Ile Leu Ile Arg Lys Gly Gln Pro Val Gly Tyr
 385 390 395 400
 Phe Ile Cys Ala Thr His Leu Asn Glu Lys Val Tyr Gly Asn Asp Ser
 405 410 415
 His Val Phe Arg Pro Glu Arg Trp Ala Ala Leu Glu Gly Lys Ser Leu
 420 425 430
 Gly Trp Ser Tyr Leu Pro Phe Asn Gly Gly Pro Arg Ser Cys Leu Gly
 435 440 445
 Gln Gln Phe Ala Ile Leu Glu Ala Ser Tyr Val Leu Ala Arg Leu Thr
 450 455 460
 Gln Cys Tyr Thr Thr Ile Gln Leu Arg Thr Thr Glu Tyr Pro Pro Lys
 465 470 475 480
 Lys Leu Val His Leu Thr Met Ser Leu Leu Asn Gly Val Tyr Ile Arg
 485 490 495
 Thr Arg Thr

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1712 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

```

GGTACCGAGC TCACGAGTTT TGGGATTTTC GAGTTTGGAT TGTTTCCTTT GTTGATTGAA      60
TTGACGAAAC CAGAGGTTTT CAAGACAGAT AAGATTGGGT TTATCAAAAC GCAGTTTGAA      120
ATATTCCAGT TGGTTTCCAA GATATCTTGA AGAAGATTGA CGATTTGAAA TTTGAAGAAG      180
TGGAGAAGAT CTGGTTTGA TTGTTGGAGA ATTTCAAGAA TCTCAAGATT TACTCTAACG      240
ACGGGTACAA CGAGAATTGT ATTGAATTGA TCAAGAACAT GATCTTGGTG TTACAGAACA      300
TCAAGTTCTT GGACGAGACT GAGAATGCCA CAGATATACA AGGCGTCATG TGATAAAATG      360
GATGAGATTT ATCCCACAAT TGAAGAAAGA GTTTATGGAA AGTGGTCAAC CAGAAGCTAA      420
ACAGGAAGAA GCAAACGAAG AGGTGAAACA AGAAGAAGAA GGTAATAAAG TATTTTGTAT      480
TATATAACAA ACAAAGTAAG GAATACAGAT TTATACAATA AATTGCCATA CTAGTCACGT      540
GAGATATCTC ATCCATTCCC CAACTCCCAA GAAAAAAGAA AAGTGAAAAA AAAAATCAAA      600
CCCAAAGATC AACCTCCCA TCATCATCGT CATCAAACCC CCAGCTCAAT TCGCAATGGT      660
TAGCACAAAA ACATACACAG AAAGGGCATC AGCACACCCC TCCAAGGTTG CCCAACGTTT      720
ATTCCGCTTA ATGGAGTCCA AAAAGACCAA CCTCTGCGCC TCGATCGACG TGACCACAAC      780
CGCCGAGTTC CTTTCGCTCA TCGACAAGCT CGGTCCCCAC ATCTGTCTCG TGAAGACGCA      840
CATCGATATC ATCTCAGACT TCAGCTACGA GGGCACGATT GAGCCGTTGC TTGTGCTTGC      900
AGAGCGCCAC GGGTCTTGA TATTCGAGGA CAGGAAGTTT GCTGATATCG GAAACACCGT      960
GATGTTGCAG TACACCTCGG GGGTATACCG GATCGCGGCG TGGAGTGACA TCACGAACGC      1020
GCACGGAGTG ACTGGGAAGG GCCTCGTTGA AGGGTTGAAA CGCGGTGCGG AGGGGGTAGA      1080
AAAGGAAAGG GCGGTGTTGA TGTGCGGGA GTTGTCTGAGT AAAGGCTCGT TGGCGCATGG      1140
TGAATATACC CGTGAGACGA TCGAGATTGC GAAGAGTGAT CGGGAGTTCG TGATTGGGTT      1200
CAGCGCGCAG CGGGACATGG GGGGTAGAGA AGAAGGGTTT GATTGGATCA TCATGACGCC      1260
TGGTGTGGGG TTGGATGATA AAGGCGATGC GTTGGGCCAG CAGTATAGGA CTGTTGATGA      1320
GGTGGTCTG ACTGGTACCG ATGTGATTAT TGTCTGGAGA GGGTTGTTTG GAAAAGGAAG      1380
AGACCCTGAG GTGGAGGGAA AGAGATACAG GGATGCTGGA TGAAGGCAT ACTGAAGAG      1440
AACTGGTCAG TTAGAATAAA TATTGTAATA AATAGGTCTA TATACATACA CTAAGCTTCT      1500
AGGACGTCAT TGTAAGTCTC GAAGTTGTCT GCTAGTTTAG TTCTCATGAT TTCGAAAACC      1560

```

AATAACGCAA TGGATGTAGC AGGGATGGTG GTTAGTGCGT TCCTGACAAA CCCAGAGTAC 1620
 GCCGCCTCAA ACCACGTCAC ATTCGCCCTT TGCTTCATCC GCATCACTTG CTTGAAGGTA 1680
 TCCACGTACG AGTTGTAATA CACCTTGAAG AA 1712

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 267 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

Met	Val	Ser	Thr	Lys	Thr	Tyr	Thr	Glu	Arg	Ala	Ser	Ala	His	Pro	Ser	1	5	10	15
Lys	Val	Ala	Gln	Arg	Leu	Phe	Arg	Leu	Met	Glu	Ser	Lys	Lys	Thr	Asn	20	25	30	
Leu	Cys	Ala	Ser	Ile	Asp	Val	Thr	Thr	Ala	Glu	Phe	Leu	Ser	Leu		35	40	45	
Ile	Asp	Lys	Leu	Gly	Pro	His	Ile	Cys	Leu	Val	Lys	Thr	His	Ile	Asp	50	55	60	
Ile	Ile	Ser	Asp	Phe	Ser	Tyr	Glu	Gly	Thr	Ile	Glu	Pro	Leu	Leu	Val	65	70	75	80
Leu	Ala	Glu	Arg	His	Gly	Phe	Leu	Ile	Phe	Glu	Asp	Arg	Lys	Phe	Ala	85	90	95	
Asp	Ile	Gly	Asn	Thr	Val	Met	Leu	Gln	Tyr	Thr	Ser	Gly	Val	Tyr	Arg	100	105	110	
Ile	Ala	Ala	Trp	Ser	Asp	Ile	Thr	Asn	Ala	His	Gly	Val	Thr	Gly	Lys	115	120	125	
Gly	Val	Val	Glu	Gly	Leu	Lys	Arg	Gly	Ala	Glu	Gly	Val	Glu	Lys	Glu	130	135	140	
Arg	Gly	Val	Leu	Met	Leu	Ala	Glu	Leu	Ser	Ser	Lys	Gly	Ser	Leu	Ala	145	150	155	160
His	Gly	Glu	Tyr	Thr	Arg	Glu	Thr	Ile	Glu	Ile	Ala	Lys	Ser	Asp	Arg	165	170	175	
Glu	Phe	Val	Ile	Gly	Phe	Ile	Ala	Gln	Arg	Asp	Met	Gly	Gly	Arg	Glu	180	185	190	
Glu	Gly	Phe	Asp	Trp	Ile	Ile	Met	Thr	Pro	Gly	Val	Gly	Leu	Asp	Asp	195	200	205	
Lys	Gly	Asp	Ala	Leu	Gly	Gln	Gln	Tyr	Arg	Thr	Val	Asp	Glu	Val	Val	210	215	220	
Leu	Thr	Gly	Thr	Asp	Val	Ile	Ile	Val	Gly	Arg	Gly	Leu	Phe	Gly	Lys	225	230	235	240
Gly	Arg	Asp	Pro	Glu	Val	Glu	Gly	Lys	Arg	Tyr	Arg	Asp	Ala	Gly	Trp	245	250	255	
Lys	Ala	Tyr	Leu	Lys	Arg	Thr	Gly	Gln	Leu	Glu						260	265		

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 473 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

GTCAAAGCAA	ATTGTTGGCC	CAAGCAGACT	CTTGGACCAC	CGTTGAATGG	AACATAAGCC	60
CAGCCCAACT	TCTTAGTAGA	TGGTTCAAAC	CATCTTTCTG	GTCTGAAGTC	GTTAGCGTCC	120
TTACCGTAGT	ATTCTTCCAA	ACGGTGGGTC	TTGTAGACAA	CGTAAGCAAC	AGTGGAGCCT	180
TTAGGAATGT	AGATTGGGTC	GGTACCGTTA	GCACCACCAC	CTCTTGCCAA	AGTGGTGTCT	240
CTGGTGGCGG	TTCTAAAGTT	GACAGGAACA	GATGGGTACA	TACGCAAGGT	TTCGTTAAGG	300
ATAGCCTTCA	AGTATTCACA	TCTCTTCAAG	GCTTCGAAAG	TAATTCTTTC	AACGCGGGAG	360
TCTTCACCAA	CACCAAAGTT	AACTTCGATT	TCTTCTCTCA	ACTTGGACCA	CATCTCTGGG	420
TGTCTAGCCA	ATTCAAACAA	AGCAAAGGAC	AACAAACCCG	CGGTGGTGTC	TCT	473

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
CCTTAATTAA GAGGTCGTTG GTTGAGTTTT C 31
- (2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
CCTTAATTAA TTGATAATGA CGTTGCGGG 29
- (2) INFORMATION FOR SEQ ID NO:9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
AGGCGCGCCG GAGTCCAAA AGACCAACCT CTG 33
- (2) INFORMATION FOR SEQ ID NO:10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
CCTTAATTAA TACGTGGATA CCTTCAAGCA AGTG 34
- (2) INFORMATION FOR SEQ ID NO:11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
CCTTAATTAA GCTCACGAGT TTTGGGATTT TCGAG 35
- (2) INFORMATION FOR SEQ ID NO:12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
GGGTTTAAAC CGCAGAGGTT GGTCTTTTGT GACTC 35